

THE STRUCTURE OF THE UNIVERSE
AN INTRODUCTION TO COSMOLOGY

HUTCHINSON'S UNIVERSITY LIBRARY
MATHEMATICAL AND PHYSICAL SCIENCES

EDITOR:

PROFESSOR E. A. MILNE

M.B.E., F.R.S., M.A., M.Sc., F.R.A.S., F.R.MET.SOC., F.R.INST.P.

*Rouse Ball Professor of Mathematics
in the University of Oxford*

THE
STRUCTURE OF THE UNIVERSE
AN INTRODUCTION TO COSMOLOGY

by

G. J. WHITROW

M.A., D.PHIL. (OXON), F.R.A.S.

LECTURER IN MATHEMATICS AT THE
IMPERIAL COLLEGE OF SCIENCE AND
TECHNOLOGY, FORMERLY LECTURER
OF CHRIST CHURCH, OXFORD

HUTCHINSON'S UNIVERSITY LIBRARY
11 Stratford Place, London, W.1
New York *Melbourne* *Sydney* *Cape Town*

THIS VOLUME IS NUMBER 29 IN
HUTCHINSON'S UNIVERSITY LIBRARY

*Printed in Great Britain
by The Anchor Press, Ltd.,
Tiptree, Essex*

CONTENTS

<i>Chapter I</i>	The Depths of the Universe (I)	<i>Page</i> 7
II	The Depths of the Universe (II)	25
III	Space and Time	41
IV	Relativity	57
V	World-Models (I)	75
VI	World-Models (II)	92
VII	The Age of the Universe	116
VIII	The Structure of the Nebulae	127
IX	Cosmology and the <i>A Priori</i>	148
	Select Bibliography	161
	Index	164

ERRATA

Page 97:

First paragraph: For N read $\frac{1}{2}N$ throughout.

Second paragraph: For *protons* read *particles*.

Page 98:

In equation (3) and subsequent equation *delete* the factor $\frac{1}{2}$.

In equation (4) *delete* the factor 4.

CHAPTER I

THE DEPTHS OF THE UNIVERSE (I)

“LET man then contemplate the whole of nature in her full and grand mystery, and turn his vision from the low objects which surround him. Let him gaze on that brilliant light, set like an eternal lamp to illumine the universe; let the earth appear to him a point in comparison with the vast circle described by the sun, and let him wonder at the fact that this vast circle is itself but a very fine point in comparison with that described by the stars in their revolution round the firmament. But if our view be arrested there let our imagination pass beyond; it will sooner exhaust the power of conception than nature that of supplying material for conception. The whole visible world is only an imperceptible atom in the ample bosom of nature. No idea approaches it. We may enlarge our conceptions beyond all imaginable space; we only produce atoms in comparison with the reality of things. It is an infinite sphere, the centre of which is everywhere, the circumference nowhere.” So wrote the great French mathematician Pascal three hundred years ago.

For centuries innumerable men had imagined the whole physical universe to be enclosed in a *finite* sphere. The sun and moon and the five planets were the principal heavenly bodies, to each of which was ascribed a special domain in which it traced its path. By an elaborate and ingenious theory their observed motions were adequately described in terms of epicycles, or circles whose centres rotate on other circles. In the background was the spinning sphere of stars, for these were all supposed to be at the same distance from the earth, the hub of the universe.

When Pascal wrote, faith in this simple picture of the world was being undermined by one of the most momentous revolutions which have occurred in the history of thought. Much had happened in the previous hundred years to cause men to question their former ideas. The old scholastic teach-

ings of the universities had degenerated into futile logomachies, and the more alert minds had turned away from the fixed mediaeval order of nature to humanism with its less rigid and more relativistic point of view. In the world of trade and travel men were also becoming conscious of wider horizons. After the famous voyages of Henry the Navigator, Columbus and Magellan, it no longer seemed that the only centre of importance in the world was Europe. Furthermore, the Reformation had shaken the old belief that there was a unique religious centre. The rise of Protestantism, with its emphasis on personal faith, was perhaps the most powerful disruptive influence, but also in art and letters there was a tendency to nationalism which favoured the idea that there was no one fixed centre of the world.

Into this explosive mental atmosphere an obscure Polish cleric launched a thunderbolt. In his *De Revolutionibus Orbium Caelestium*, published in 1543, Copernicus suggested that the earth was not the unique centre of the astronomical universe, any more than Jerusalem or Rome was the centre of the habitable region of the earth. He argued that the mathematical description of the observed motions of the heavenly bodies could be simplified if the earth was regarded no longer as fixed and if all motion was depicted as being relative to the sphere of the fixed stars, at the centre of which he placed the sun. His argument was mainly mathematical, but he pointed out that, if it were objected that a moving earth would fly to pieces, then *a fortiori* a moving crystalline sphere of stars, rotating at a far greater rate, would be even more likely to disintegrate. The daily rotation of the heavens was thus transferred to the earth.

The next advance was made by Thomas Digges in 1576, and also by the famous Giordano Bruno, whose restless imagination could no longer be content with the idea that the stars outside the solar system were enclosed in a shell concentric with either the earth or the sun. A serious objection raised against the Copernican system was that, if the earth rotated about the sun, then in the course of the year there should be a distinct annual change in the apparent positions of the stars relative to the earth. As such an annual motion or parallax, as it is usually called, was not observed, the only answer which the

upholders of the new theory could offer was that the stars were immensely distant compared with the planets.

We have seen that, in referring to the stellar realm beyond the planets, Pascal remarked, "if our view be arrested there let our imagination pass beyond". Despite the invention of the telescope and its application to astronomical observation early in the seventeenth century by Galileo, the region of the stars was not explored until long after Pascal wrote. The stars were seen merely as a framework bounding the region in which the astronomer could conduct his investigations. Whether the stars were all at the same distance, or whether they were scattered throughout infinite space, or whether they formed a finite system of vast but limited depth, were questions that could not be answered until towards the end of the eighteenth century. Until then stellar astronomy was a field left to the unaided imagination.

Due to the genius and labours of Newton (1642-1727), almost all the problems presented by the motions of the planets had been mastered. Newton had shown for all time that these motions could be completely accounted for if it were assumed that the same laws of nature, and in particular gravity, operated in the celestial realm as well as in the terrestrial. Although the old Aristotelian distinction between the corrupt earth and the incorruptible heavens was thus finally abandoned, the stellar realm still lay beyond the range of scientific investigation. The natural step, taken by Digges and Bruno, of likening the stars to the sun and scattering them throughout space was still only a step of the imagination.

Copernicus made his revolutionary suggestion on purely theoretical grounds. The invention of the telescope came after, and not before, the publication of his great work. This is a fact on which we may well ponder, for it shows that insight into natural phenomena has come as much from the genius of the theoretical as from that of the experimental and observational investigator.

However, the next significant advance in the growth of our conception of the universe was due to one of the greatest observational astronomers of all time, William Herschel (1738-1822). By profession a musician, Herschel came to this

country from Hanover, where he had been a bandsman in the Foot Guards. He made rapid progress in the musical profession, and in 1766 was appointed organist of the Octagon Chapel at Bath. This post he held for sixteen years, but all his spare time was devoted to optics and astronomy. Too poor to afford a telescope but, to quote his own words, "resolved to take nothing upon trust but to see with my own eyes all that other men had seen before", he finally decided to make a telescopic mirror for himself. By the spring of 1774, after two hundred failures, he constructed his first reflecting telescope, and within seven years made one of the most sensational discoveries in the history of astronomy, the first planetary discovery since prehistoric times, Uranus. Herschel's success was due to his ability to construct better telescopes than had ever been made before, to the keenness of his vision ("seeing is in some respects an art which must be learnt") and above all to his genius.

The greatest enterprise of Herschel's life was his attempt to map the stellar regions beyond the solar system. Significantly, the first problem which he attacked was that of stellar parallax. This is the problem of determining the apparent angular displacement in the sky of a given star in the course of the earth's journey from one point of its orbit round the sun to the diametrically opposite point. The discovery of Uranus was a chance by-product of this investigation. As Uranus is about twice as far from the sun as Saturn, the next nearest planet, Herschel's discovery incidentally doubled the known extent of the solar system; but, from Herschel's point of view, its most important consequence was that he came under royal patronage. George III created for him the post of King's Astronomer with a salary of £200 a year. This enabled him to abandon music and devote his time and energy entirely to astronomy and above all to his great project of a systematic survey of the heavens.

Herschel was the pioneer of modern statistical astronomy. He instituted 'star gauges' or counts of the number of stars visible in his telescope in different parts of the sky, and two great problems emerged therefrom. One was the problem of determining the sun's own motion among the stars, and the other was the investigation of the shape and structure of the

Milky Way, or Galaxy, which Galileo had found early in the seventeenth century to be composed of a vast number of faint stars.

In 1718 the great British astronomer Halley discovered that the positions of three bright stars—Sirius, Arcturus and Aldebaran—differed appreciably from their positions as recorded in the catalogue compiled by Hipparchus nearly two thousand years before, due allowance being made for the effect of the precession of the equinoxes, or top-like motion of the earth's axis of rotation. The obvious explanation of this phenomenon was that the three stars had moved at right angles to the line of sight against the general background of the other stars. The first breach was thus made in the age-old conception that the stars were fixed, but no further progress was made before Herschel. His earliest paper in 1783 is largely devoted to a philosophical attempt to prove that, as the sun is a star, it too should move against the background of the stars as a whole.

"If the proper motion of the stars in general be at once admitted," asked Herschel, "who can refuse to allow that our Sun, with its planets and comets, is no less liable to such a general agitation as we find to obtain among all the rest of the celestial bodies?" He argued that the sun's motion could only be detected by a general drift of the stars in a contrary direction, those stars ahead of the sun appearing to disperse, while those behind tend to draw together. The determination of the solar motion, however, is complicated by the fact that the stars themselves have their own proper motions. Thus the observed motion of each star would have to be decomposed into two component factors, the star's own motion and the reflection of the sun's motion in the opposite direction.

In his first paper Herschel tentatively fixed the direction of the sun's motion as being towards a point in the constellation Hercules, near the star Lambda Herculis. Moreover, although very few proper motions were known with accuracy and no one had yet succeeded in measuring the distance of a single star, in his second paper in 1806 Herschel advanced an estimate for the amount of the solar motion, which he expressed in terms of the then unknown distance of Sirius. Using the now accepted value for the latter, his result was fourteen kilometres per

second as compared with the modern value of twenty kilo-

The other main problem which engaged Herschel's attention was the general distribution of the stars. In two papers, presented to the Royal Society in 1785, he put forward the first definite model of the stellar system. He determined the relative distances of stars in 3,400 regions of known position on the assumption that all stars were of the same absolute brightness, so that apparent brightness became his criterion of relative distance. We now know that this is an extremely poor criterion for individual stars; statistically it is more satisfactory, and it was of course in this way that Herschel used it and thereby came to the conclusion that there are a finite number of stars, according to his data from 75 to 100 million. These are grouped in a bun-shaped system, the appearance of the Milky Way in the sky being an optical effect due to this shape. He estimated the diameter of the Galaxy, as we now call the system, to be 850 times and the thickness to be 155 times the average distance of a first magnitude star, such as Sirius. He believed that the sun was placed near but not quite at the centre, a result which more recent research has since modified considerably. On the other hand, his belief that the sun is very near the central plane, because the Milky Way appears to divide the sky into two almost exactly equal parts, has been entirely confirmed.

This conception of a stellar system isolated in space appears to have been first put forward as a possible hypothesis by an English instrument maker, Thomas Wright, in 1750. But his speculations did not rest there. As Hubble has so pertinently remarked, the few speculations on the general structure of the universe which have survived the results of observational exploration have all been based on the principle of uniformity—that, in a general sense, any sample of the universe is much like any other sample. In accord with this principle, Wright found the conception of a finite system of stars isolated in infinite Euclidean space unsatisfactory. As the idea of finite non-Euclidean space had not then been invented, he suggested that there exist other systems similar to the Galaxy, just as the stars are other bodies similar to the sun. Five years later the

great German philosopher Immanuel Kant developed this conception further. Some of Kant's remarks have been cited by Hubble, in a rather free translation, as an excellent example of reasonable speculation based on the principle of uniformity.

"Let us imagine a system of stars gathered together in a common plane, like those of the Milky Way, but situated so far away from us that even with the telescope we cannot distinguish the stars composing it; let us assume that its distance, compared to that separating us from the stars of the Milky Way, is in the same proportion as the distance of the Milky Way is to the distance from the earth to the sun; such a stellar world will appear to the observer, who contemplates it at so enormous a distance, only as a feeble spot feebly illumined and subtending a very small angle; its shape will be circular, if its plane is perpendicular to the line of sight, elliptical, if it is seen obliquely. The faintness of its light, its form, and its appreciable diameter will obviously distinguish such a phenomenon from the isolated stars around it.

"We do not need to seek far in the observations of the astronomers to meet with such phenomena. They have been seen by various observers, who have wondered at their strange appearance, have speculated upon them, and have suggested sometimes the most amazing explanations, sometimes theories which were more rational, but which had no more foundation than the former. We refer to the nebulae, or, more precisely, to a particular kind of celestial body which M. de Maupertuis describes as follows: 'These are small luminous patches, only slightly more brilliant than the dark background of the sky; they have this in common, that their shapes are more or less open ellipses; and their light is far more feeble than that of any other objects to be perceived in the heavens. . . .'

"It is much more natural and reasonable to assume that a nebula is not a unique and solitary sun, but a system of numerous suns, which appear crowded, because of their distance, into a space so limited that their light, which would be imperceptible were each of them isolated, suffices, owing to their enormous numbers, to give a pale and uniform lustre. Their analogy with our own system of stars; their form, which is precisely what it should be according to our theory; the faintness of their

light, which denotes an infinite distance; all are in admirable accord and lead us to consider these elliptical spots as systems of the same order as our own—in a word, to be Milky Ways similar to the one whose constitution we have explained. And if these hypotheses, in which analogy and observation consistently lend mutual support, have the same merit as formal demonstrations, we must consider the existence of such systems as demonstrated. . . .

“We see that scattered throughout space out to infinite distances there exist similar systems of stars, and that creation, in the whole extent of its infinite grandeur, is everywhere organized into systems whose members are in relation with one another. . . . A vast field lies open to discoveries, and observation alone will give the key.”

These remarkable speculations found their first substantial observational support in the investigations of Herschel, who made a systematic search for nebulae. In 1783 Messier catalogued 103 such objects, which are still known by the numbers he assigned to them. For example, the famous nebula in Andromeda is M31. In striking confirmation of Kant’s hypothesis, Herschel found that a number of these ‘nebulae’ were resolved into stars by his telescope. These ‘nebulae’ are in fact now called ‘globular clusters’, forming parts of, or adjuncts to, the Galaxy rather than totally independent and comparable systems. Others, however, appeared to be just beyond his telescope’s power of resolution. On calculating the minimum distance at which they must therefore be, Herschel concluded that “the stupendous sidereal system we inhabit, consisting of many millions of stars, is, in all probability, a detached nebula. Among the great number of nebulae which I have now already seen, amounting to more than 900, there are many which in all probability are equally extensive with that which we inhabit.”

The problem of confirming such speculations was laborious and difficult. Indeed, in later years Herschel modified his views considerably. In 1811 he wrote: “I must freely confess that by continuing my sweep of the heavens my opinion of the arrangement of the stars and their magnitudes, and of some other particulars, has undergone a gradual change. . . . We may also

have surmised nebulae to be no other than clusters of stars disguised by their very great distance, but a longer experience and better acquaintance with the nature of nebulae will not allow a general admission of such a principle."

In his first papers in 1785 Herschel had assumed an approximately equal scattering of stars in the Galaxy, for although "in all probability there may not be two or three of them in the heavens, whose mutual distance shall be equal to that of any other two given stars, but it should be considered that when we take all the stars collectively there will be a mean distance which may be assumed as the general one". However, by 1802 he fully recognized that "this immense starry aggregation is by no means uniform. The stars of which it is composed are very unequally scattered, and show evident marks of clustering together into many separate allotments." He thus came to regard the Galaxy as a system consisting of many local groups and clusters, and then to question the complete universality of his original view that all nebulae are distant 'island universes' comparable to our own galactic system. By 1811 Herschel had foreseen the modern division of nebulae into two classes of objects, the one consisting of external galaxies and the other of diffuse matter within our own galaxy, for example the nebulous mass in Orion, which he concluded was comparatively close to us, in agreement with present-day opinion.

The problem of the nature and status of the nebulae was, in fact, one of the most difficult which Herschel tackled, because he had no direct knowledge of the distances of a single object outside the solar system. In 1785, supposing them all to be unresolved aggregations of stars, he thrust them all outside the Galaxy, thus imagining he had pierced through the boundaries of the Milky Way, but by 1817 he admitted that "the utmost stretch of the space-penetrating power of the 20-foot telescope could not fathom the profundity of the Milky Way". Besides the existence of unresolved nebulae, the fact that with a 40-foot telescope he saw more stars than with a 20-foot telescope made it seem improbable that he had reached the confines of the universe; a background remained. Thus, although he still maintained his view that some of the distant nebulae were independent stellar systems, the evidence was

most inconclusive. Consequently, as Eddington has remarked, "the status of the nebulae has undergone remarkable vicissitudes". In 1864 Sir William Huggins found that the light from the Orion nebula and from some others was similar to that emanating from a glowing mass of gas. This made it impossible to regard them as aggregations of stars, and so the observational evidence appeared to favour the view that all unresolved nebulae were masses of glowing gas within the Galaxy.

In addition to his vast pioneer work on the general structure of the sidereal universe, one further discovery by Herschel is of outstanding cosmological significance. For, just as by his systematic methods of star gauging he discovered that the speculations on the distribution of matter which had been based on the principle of uniformity were in essence correct, so by observing double stars he was the first to show that the laws of Nature holding outside the solar system conformed with those prevailing within. Newton's intellectual masterpiece was the discovery that planetary and terrestrial motion could be brought within the scope of the same law of gravitation; likewise, Herschel's greatest achievement was his brilliant demonstration that the same law governed the motion of stars. In 1802 he discovered evidence of orbital rotation in a number of double stars, and in 1804 he concluded beyond doubt that "many doubles must be allowed to be real binary combinations of two stars intimately held together by the bond of mutual attraction".

Well merited are the words of Sir William Herschel's epitaph in Upton Church, *Coelorum perrupit claustra*, for he broke through the barrier of the heavens and "looked further into space than any human being ever did before him". As Sir Oliver Lodge has reminded us, before Herschel's discoveries "the stars had only been observed for nautical and practical purposes . . . they had been treated as a clock or piece of dead mechanism, and as fixed points of reference. All the energies of the astronomers had gone towards the solar system. It was the planets that had been observed. Tycho had observed and tabulated their positions. Kepler had found out some laws of their motion. Galileo had discovered their peculiarities and attendants, Newton and Laplace had perceived every detail of their

laws. . . . But for the stars—the old Ptolemaic system might still have been true. They might still be mere dots in a vast crystalline sphere, all set at about one distance, and subservient to the uses of the earth . . . Herschel changed all this. Instead of sameness, he found variety; instead of uniformity of distance, limitless and utterly limitless fields and boundless distances; instead of rest and quiescence, motion and activity; instead of stagnation, life."

One obvious criticism of Herschel's work was that it was based purely on data obtainable in England, so that all stars round the south celestial pole to within twenty degrees of the celestial equator were left out of account. To remedy this omission, his son, Sir John Herschel, transported a 20-foot telescope to South Africa in 1834. Within four years he succeeded in corroborating the general views of his father, although it appeared that the sun was not so centrally situated within the Galaxy as he had imagined, being displaced somewhat towards the great star clouds forming its southern boundaries. The most remarkable result, however, of the younger Herschel's survey was the discovery that the nebulae, over seventeen hundred of which were catalogued and classified, tended to avoid the girdle of the Milky Way and to congregate most thickly towards the galactic poles, as the regions of the sky farthest away from this girdle are called.

The main criticism, however, of Herschel's pioneer investigations was that he had failed to discover the yard-stick of the universe, for he had no means other than plausible hypotheses for estimating the distance of a single star. As his whole conception of the sidereal universe turned on the assumption that the stars were at different distances, clearly the next step in the march of investigation must be the determination of these distances.

Although the classical theoretical foundation of distance measurement in physics is the 'rigid rod', nearly all distances in surveying, whether terrestrial or celestial, are made to depend on the properties of light. The two simplest properties so employed are the principle of propagation in straight lines and the principle that the intensity of light diminishes inversely as the square of the distance.

The former is the basis of trigonometrical surveying. It was used in antiquity by Aristarchus to make the first scientific attempt to measure the size and distance of the sun and moon by considering the phenomena to be observed at an eclipse of the moon and also when the moon was half full. Although his methods were accurate in principle, the results he obtained were much too small, owing to the technical difficulty of measuring the necessary angles with sufficient accuracy. He arrived, however, at the revolutionary conclusion that the sun was larger than the earth, and that in modern measurement its distance was over four million miles, about one-twentieth of the true figure.

The latter property was used in the second half of the seventeenth century by Newton and by Huygens to estimate stellar distances. To obtain a first crude approximation, they assumed the sun to be a star and also that all stars had the same absolute magnitude of brightness. Thus the apparent faintness of the stars was used to obtain the general order of magnitude of their distances. For example, Huygens compared the light of Sirius with that of the reflection from a small sphere of glass of a small portion of the sun's disc and found the star to be twenty-eight thousand times more distant than the sun. This figure is about twenty times too small, partly because Sirius has since been found to be about twenty-five times as bright as the sun and partly because Huygens' measurements were faulty. Similarly, Newton concluded that the average distance of the stars of the first magnitude was about thirty times Huygens' estimate for Sirius.

These first crude approximations revealed the enormous scale of the stellar universe compared with that of the solar system and also showed the inadequacy of existing technique to estimate stellar distances trigonometrically; for this method depends on measuring the parallax or apparent swinging motion of a star resulting from the earth's orbital motion. Nevertheless, during the eighteenth century numerous attempts were made to measure this phenomenon, but in each case without success.

The key to the problem was the discovery, already noted, of stellar motion by Halley in 1718. More than a century later a

German astronomer, Bessel, director of the Königsberg observatory, noted that a faint object, numbered 61 Cygni, *i.e.* 61 in the constellation of the Swan, had the fastest movement of any star then known, an interval of 300 years apparently sufficing to carry it across the sky through a distance equal to the diameter of the moon. He concluded that such rapid angular motion must be due to relative proximity, despite the object's apparent faintness.

Until then all attacks on the problem of stellar distance had been confined to the apparently brightest stars, but in December 1838 Bessel published his first results on the distance of this fainter star, making full use of the latest developments in observational technique and in the principles of optics. In the previous year Struve, who afterwards became the first director of the famous Pulkowa observatory near St. Petersburg, had detected the parallactic motion of Vega, but his results were not very consistent. Bessel's results, however, were; and his estimated distance of 61 Cygni as about 50 million million miles, or over half a million times as far away as the sun, is now recognized to be the first reliable measurement of a star's distance, differing from the now accepted value by only a few per cent.

The scale of the stellar system is clearly one to which terrestrial units, *e.g.* the mile, are inappropriate. In practice, the two most convenient units of distance in stellar astronomy are the light-year and the parsec. The light-year is the distance travelled by light in *vacuo* in one year at the rate of about 186,000 miles or 300,000 kilometres per second, *i.e.* about six million million miles. The parsec is the distance at which a star would exhibit a parallax of one second of arc (one thirty-six-hundredth of a degree). It is about three and a quarter light-years. Thus 61 Cygni is now estimated to be 11.1 light-years distant, and the nearest star about 4.3 light-years.

Herschel's memoirs on the structure of the universe were studied in great detail by Struve, whose own researches led him to regard the whole system of stars, clusters and nebulae as of finite thickness but of infinite extension in the galactic plane. Herschel's original idea of a finite galaxy surrounded by other stellar systems or island universes thus fell further into dis-

favour. The discovery, already mentioned, that the unresolved nebulae tended to avoid the galactic plane appeared to be further evidence against the view that they were *independent* stellar systems, and during the remainder of the nineteenth and the first decade of the twentieth century the focal point of interest in cosmology was the general structure of the Galaxy.

The remarkable degree of success achieved by Herschel in elucidating the structure of the sidereal system, despite his ignorance of data concerning individual stars, was primarily due to his skill in devising and using statistical methods. These methods were greatly extended and elaborated between 1884 and 1911 by the German astronomer Seeliger, who introduced the fruitful idea of a frequency distribution in the intrinsic brightness of the stars. This was a great advance over Herschel's assumption of uniform intrinsic brightness, for it came to be widely recognized that individual stars vary remarkably in their luminosity. Seeliger's idea was to count the number of stars between different ranges of apparent brightness, and not merely the total number, visible in different parts of the sky. Thus it was possible to determine the relative rate at which the stars fell off in number at increasing distances from the centre of the sidereal system. Seeliger's final model was a watch-shaped system similar to Herschel's, the star density decreasing with distance from the sun. The estimated diameter was 7,250 parsecs or 23,000 light-years, about four times the figure assigned by Herschel, and the thickness nearly 6,000 light-years.

The statistical analysis of stellar distribution was refined by several astronomers and above all by the great Dutch astronomer Kapteyn, who spent twelve years on the stupendous task of measuring and recording the positions and brightnesses of nearly half a million stars photographed in the southern hemisphere. This achievement appears all the more remarkable when we realize that the work was carried out practically single-handed in two small rooms borrowed from the physiological department of the University of Groningen, although at one time he had the help of a few convicts loaned to him from a Dutch prison.

The most important advance in method made by Kapteyn

was his scheme of 'selected areas'. As the number of stars in our Galaxy is probably of the order of a hundred thousand million it is clearly impossible to count more than a minute fraction. To obtain reliable results it is therefore of the utmost importance to select our samples fairly. As most of the data related to the brighter and hence more easily observed stars, Kapteyn proposed that 252 small regions of the sky distributed more or less evenly over the heavens be chosen and observatories all over the world be asked to co-operate in recording data concerning every star from the brightest to the faintest that could be observed in these regions. This scheme is still far from complete, but by fixing the limits of the Galaxy where the star density falls off to one-hundredth of its value near the sun, Kapteyn obtained a distribution similar in shape to those of Herschel and Seeliger but with a diameter of 17,000 parsecs or about nine times that assigned by Herschel, and a thickness of 3,500 parsecs.

One feature of the Kapteyn system calls for immediate comment. In his model of the Galaxy, as in those of Herschel and Seeliger, the sun still occupies a central position. The validity of this conception was open to question. The whole trend of astronomy since Copernicus had been against the previously prevailing idea that Man occupied a preferential position in the physical universe. As the earth had been dethroned from its central position in the solar system, it was also felt that there was no reason why the sun should be central in the Galaxy. Indeed, in view of the enormous number of stars, the probability was overwhelmingly against this concept.

Furthermore, it was realized that a very small amount of absorption of light, for example by diffuse material in interstellar space, would completely upset the calculated density distribution, and thus the results so laboriously achieved could only be regarded as provisional. Indeed, recent investigations have led us to believe that we live in a kind of celestial fog which is densest in and near the plane of the Milky Way. The effect of this obscuring matter is believed to be so great that light is halved in intensity after travelling about a thousand parsecs. It follows that the majority of the more distant objects in the Galaxy cannot be seen, and we are deceived into imagining that

we are more centrally situated than in fact we are. Moreover, it is now realized that the distances formerly assigned to faint objects were probably too large, so that previous estimates of the dimensions of the sidereal system were excessive.

However, the principal cosmological significance of the existence of interstellar obscuring matter is that a new explanation is suggested for the fact that the nebulae appear to avoid the plane of the Milky Way. None are to be seen there, not because there are none in those directions, but because the general obscuration of light prevents our observing them. Thus the hypothesis that the nebulae must be satellite systems to the Galaxy is greatly weakened, because the mere fact that they shun the central plane can no longer be taken to indicate that they are influenced by it.

The crucial factor, however, in deciding between these rival speculations, and also in dethroning the sun from its central position in the Kapteyn universe, was the discovery of a remarkable new astronomical yard-stick. Again our knowledge of the universe has been extended by a new application of the principle of the diminution of light intensity with the square of the distance. In 1912 Miss Leavitt, of the Harvard Observatory, found an empirical relation between the periods and luminosities of a certain type of variable star. These stars, known as Cepheids, after their prototype Delta Cephei, when regularly observed yield a typical and easily recognized light fluctuation curve. There is a rapid increase of light, followed by a more gradual decline and then again by the same rapid increase and slow decline as before. Near the Galaxy is a particular cluster of stars known as the Lesser Magellanic Cloud. In this cloud there are a number of Cepheid variables, and Miss Leavitt discovered that the light of the brighter Cepheids fluctuated more slowly than the light of the fainter ones. Since the Magellanic Clouds are aggregations of very faint stars, it can be assumed for all practical purposes that these stars are equally distant, so that the relation between period of fluctuation and apparent brightness can be interpreted as due to a relation between period and intrinsic brightness, or absolute magnitude.

Miss Leavitt's rule has been amply confirmed by other investigators since, as Cepheid variables occur in many parts of

the heavens. Thus, by invoking the inverse square law, her rule can be used to determine the relative distances of these stars. The principal difficulty occurs in determining the factor converting relative into absolute distances. However, the absolute distances of a few of the nearer Cepheids have been determined by measuring their parallaxes, although owing to the remoteness of these stars the uncertainty in these measurements may be at least ten per cent.

This new method of estimating stellar distances is particularly valuable because it can be applied where other methods fail owing to the remoteness of the object concerned. It was used by the American astronomer Harlow Shapley to determine the distances of the globular clusters of stars surrounding the Milky Way. About a hundred of these clusters are known and they all look much alike except for apparent size, and this is probably mainly due to differences of distance. Cepheid variables abound in them, and Shapley found that the distances of these clusters range from about 20,000 to about 200,000 light-years. At such distances the parallactic method of measuring distance would be hopelessly inapplicable; as Jeans has remarked, the parallactic orbit of a star at the distance of the furthest globular clusters is about equivalent to the size of a pin-head held at the other side of the Atlantic.

From the cosmological aspect, the most significant result of Shapley's work was the discovery that the sun is near one edge of the cluster system, the geometrical centre of which was found to be in the direction of Sagittarius, where the star clouds are richest. When the clusters were plotted on a plane perpendicular to the central plane of the Milky Way and passing through the sun and the centre of the globular system, it was found that they were symmetrically distributed. Assuming that this indicated some intimate connection between the stellar and globular systems, Shapley argued that the two were probably concentric. He thus estimated the Galaxy to have a diameter of 80,000 parsecs or about a quarter of a million light-years and a thickness of about 5,000 parsecs, the sun being at about 20,000 parsecs from the centre. Shapley's model was therefore about five times as wide as Kapteyn's.

More recent work has tended to reduce these dimensions,

which made no allowance for the effect of interstellar obscuring matter and were also based on the assumption that the effective boundary of the main stellar system is to be found in the region of the globular clusters. Thus, although the distance of the sun from the centre of the Galaxy is still somewhat uncertain, it is now thought to be approximately 10,000 parsecs, about one-half of Shapley's original estimate. The total number of stars in the Galaxy is now thought to be of the order of a hundred thousand million.

CHAPTER II

THE DEPTHS OF THE UNIVERSE (II)

OUR ability to estimate the distance of any group of stellar objects, in which Cepheid variables of known period and apparent brightness can be identified, has found its most outstanding application in the solution of the old problem of the status of the nebulae. Herschel had originally placed them all outside the Galaxy, but later came to the conclusion that some were constituent members of this system. During the nineteenth century it was found that some of the nebulae were gaseous and definitely within the boundaries of the Galaxy. Other nebulae were found to have a peculiar structure.

In 1845 Lord Rosse completed his famous reflecting telescope, having a gigantic mirror six feet in diameter. Erected in a country district of Ireland, mainly with local labour, and with none of the advantages which are provided by modern tools, it was one of the greatest *tours de force* in the history of telescopic construction. It could not compete with modern instruments in exploring the depths of space, because these regions can only be studied satisfactorily by photographic methods using very long exposures. In 1845 these had not been invented, and in any case the telescope, owing to its mechanical and optical shortcomings, was not suited for their employment.

Nevertheless, with this huge but primitive instrument Lord Rosse made one of the capital discoveries of observational astronomy. Scarcely two months after its completion it was directed to the nebula M51, near the tail of the Great Bear. Sir John Herschel had previously depicted this nebula as a split ring surrounding a bright nucleus, but the 6-foot mirror yielded a very different picture. For the first time the famous spiral structure, lavishly employed by nature in the organic world, was revealed in the heavens. Nothing like it had ever been seen by any previous observer, but Lord Rosse ultimately catalogued fourteen nebulae of this particular shape. However, so faint are these objects that few more could be discovered

until the advent of photographic methods. With improved telescopes and exposures of many hours more and more have been revealed, and even before the construction of the 100-inch telescope at Mount Wilson in 1918 it was estimated that the number of these spiral nebulae visible in the heavens must be of the order of half a million at least.

Despite their peculiar structure and enormous numbers, these nebulae too were thought to be constituent members of the Galaxy. However, as a result of observations made with the Mount Wilson telescope opinion rapidly changed. The decisive criterion was found by Hubble in 1924 when he succeeded in identifying Cepheid variables in the nearer spirals. He thus proved conclusively that the distances of these objects is of the order of a million light-years or about five times that of the globular star clusters. Hence the spirals lay definitely outside the Galaxy, and our conception of the depth of the universe was correspondingly magnified.

The history of cosmology is a history of receding backgrounds. After Hubble's capital discovery, it was at last clear that the Galaxy is not the ultimate background of the universe. Beyond lie other systems, and it is now believed that with the 100-inch telescope many millions of external or extra-galactic nebulae can be seen. The estimation of their distances was the first step in the study of the structure of this super-system of the heavens.

Cepheid variables can be identified only in a comparatively small number of the nebulae, but by estimating their distances Hubble found in 1929 that the brightest constituent stars of these nebulae were all of about the same absolute luminosity, averaging about 50,000 times that of the sun. Thus, these stars provide a second criterion for estimating the distances of those nebulae, about 150 in number, in which individual stars can be detected. Furthermore, this selection of nebulae has been shown by Hubble to provide a third statistical criterion of distance applicable when other methods fail. Most of these nebulae are found to have luminosities falling within a range of one-half to twice the average, which is about 1,700 times brighter than that of their brightest stars or about 85 million times as bright as the sun. Thus the distribution of absolute magnitude

is much more concentrated for nebulae than for stars; hence, while Herschel's assumption that the first magnitude stars are at one distance, the second magnitude stars at another and so on, is no longer used, the same principle is still being employed in studying the distribution of the nebulae. The criterion is, of course, essentially statistical, and although individual results are uncertain, mean results for large numbers of nebulae are thought to be reliable.

This method of successive calibration can be used to the very limits of the telescope. The weak link in the chain is the second, the brightest star criterion; for we cannot be sure that the collection of nebulae to which it applies constitutes a fair sample. However, assuming the general validity of this method, we estimate that the faintest nebulae which can be detected with the largest telescope yet in use, the Mount Wilson 100-inch reflector, may be about 500 million light-years away. It is an awe-inspiring thought that some of the minute specks of light which we now find darkening our photographic plates have been travelling through intervening space for about as long a period as has elapsed since the oldest terrestrial fossils were originally formed.

Within the volume of space accessible to the 100-inch telescope it is estimated that there are scattered about a hundred million nebulae of average candle-power nearly one hundred million times that of the sun, the regions between these stellar systems being comparatively free of diffuse matter. Nebulae are found both singly and in clusters, but on the grand scale Hubble claims that the distribution appears to be approximately uniform, the average distance between nebulae being of the order of 2 million light-years.

Having finally decided that the vast majority of nebulae lie outside the Galactic system, the next step was to determine their approximate size. Once the distances of the nearer nebulae, such as the great nebula in Andromeda, had been estimated to within a probable error of about ten per cent., it was a comparatively simple matter to calculate their dimensions from the images recorded on photographic plates. At first it appeared that these nebulae were several times smaller than the Galaxy. However, during recent years it has been realized that the

plates only show the regions in which stars are numerous enough to affect them to an extent visible to the unaided eye. But when the plates are studied with the aid of precise measuring apparatus, it is found that the nebulosity extends to much greater distances than previously observed. Furthermore, in the case of the Andromeda nebula for example, Hubble discovered in 1932 that it was accompanied by over a hundred objects which from their form, structure and appearance he concluded to be globular clusters similar to those accompanying the Galaxy. These clusters extended well beyond the ordinary photographic limits and about doubled the diameter of the nebula. Thus, in recent years there has been a pronounced tendency to regard this nebula as possessing a similar structure to our own Galaxy.

Moreover, just as the latest research has indicated that earlier estimates of the size of this external galaxy were probably too small, so with increasing knowledge concerning interstellar light-absorbing material it appears that Shapley's original estimate of the size of the Galaxy was probably too large, the apparent faintness of its more distant constituents being increased by the additional amount of interstellar fog to be penetrated before their light can reach the earth. Thus, it is now generally agreed that the Andromeda nebula, which for so long was thought to be much smaller than our Galaxy, is of comparable size.

When we come to compare the Andromeda nebula with other nebulae outside the Galaxy, we encounter a new complication—the question of shape. The form of our own Galaxy is not easy for us to assess owing to our internal position, but it has long been known that many of the nebulae which are now regarded as lying outside the Galaxy do not exhibit the characteristic spiral shape. Of course, this can often be attributed to their faintness, only the central portions showing up on the photographic plate; indeed, in the great majority of cases the plates reveal mere formless specks similar to the images of faint stars. The first systematic study of the shape of a group of nebulae which are all at about the same distance from us was made by Hubble in 1926. Any difference in apparent shape in such a group must be due to a difference in real shape. Hubble

examined a cluster of 167 nebulae in Virgo and discovered that only about one in two were of spiral form.

Other investigations since have tended to confirm that a considerable percentage of nebulae are definitely not spiral, but are either elliptical or else highly irregular. However, owing both to their distance and to the fact that in nearly every cluster of nebulae both forms occur, it is clear that we are justified in placing all these objects in the same category. Hence we call all these objects collectively the extra-galactic nebulae, to distinguish them from the much smaller gaseous type which is found to lie inside the Galaxy.

Another criterion which confirms this classification is that of spectral type. It is well known that light is radiated in waves of different lengths. In general, light from a given source is composite, and with the aid of a glass prism or grating can be decomposed into different colours, each colour corresponding to a definite wave-length, the deflections by the prism being least for the long red waves and greatest for the short violet waves. The whole rainbow sequence is called a spectrum, and each definite colour or relative position in the spectrum corresponds to a definite wave-length. The relative brightness of any colour indicates the relative abundance of radiation of the corresponding wave-length. Different types of radiating bodies give rise to characteristic spectra. Thus an incandescent solid yields a continuous spectrum; on the other hand, an incandescent gas radiates only isolated colours, so that its spectrum, known as an emission spectrum, consists of bright lines and dark gaps. As each gas has a spectrum of a characteristic pattern, it is possible to identify the constituent gases in a distant light source by spectrum analysis.

A star such as the sun presents a third type of spectrum, known as an absorption spectrum. The main body of the sun provides a continuous spectrum; but the surrounding solar atmosphere, although gaseous, does not furnish an emission spectrum because it is at a lower temperature. Instead, the gas absorbs from the continuous background the colours which it would normally emit. Consequently, the solar spectrum has a continuous background on which is superimposed a pattern of dark lines which are due to gaps in the continuous background.

Certain nebulae exhibit emission spectra and hence must be gaseous. These nebulae have all been found to lie within our own Galaxy. On the other hand, the extra-galactic nebulae exhibit absorption spectra characteristic of starlight and so are probably composed of stars, and possibly of dark matter too.

The sun's spectrum is dominated by the lines of hydrogen and certain metallic vapours, in particular a pair of lines in the violet region due to calcium, known as the H and K lines. It happens that the spectra of the extra-galactic nebulae are very similar to the solar spectrum, indicating that a high proportion of the light from these stellar systems must be due to stars of the same type as the sun. It is a most fortunate circumstance that, although the light from the nebulae is so faint that it can be spread only over very short spectra, it is possible to identify some of their lines and, in particular, the H and K lines of calcium.

Spectrum analysis was first applied successfully to the sun about 1814 by Fraunhofer, who established conclusively that the sun produces an absorption spectrum; the characteristic dark lines in the solar spectrum are known as Fraunhofer's lines. Forty years, however, elapsed before the general principles of spectroscopy were laid down by Kirchhoff. Nevertheless, as early as 1842 one of the key principles was enunciated by an Austrian mathematician, Christian Doppler; it is of outstanding importance in its application to the stars and nebulae. It is well known that the whistle of an engine is shriller when approaching than when receding. This is due to the fact that in the former case the sound waves are more compressed. Similarly, Doppler argued that the light waves from an approaching source are also compressed and those from a receding source correspondingly elongated, the degree of compression or elongation depending on the speed. A general shortening of wave-length would mean that the spectrum would be shifted bodily towards the violet, while a lengthening would shift it towards the red, the fractional displacement (ratio of shift to original wave-length) in each case being equal to the ratio of the speed of the source to the speed of light. Of course, in general this shift of colour will be almost imperceptible.

Nevertheless, the great British pioneer of astronomical spectroscopy, Sir William Huggins, realized that Doppler's principle could be used to determine the radial velocity of celestial bodies; that is, their velocity in the line of sight. "It would scarcely be possible," said Huggins, "to convey any true conception of the difficulties which presented themselves in this work from various instrumental causes and of the extreme care and caution which were needed to distinguish spurious instrumental shifts of a line from a true shift due to a star's motion."

However, Huggins overcame all difficulties and obtained the first crude measurements; he announced in 1868 that Sirius was receding from the solar system at twenty-nine miles per second, the correct order of magnitude. He also measured the radial velocities of other well-known stars.

Huggins' pioneer work was purely visual. The full flowering of spectroscopic astronomy came only with the application of photography at the end of the nineteenth century. Hence, exploration of the very difficult field of nebular spectra was not begun until the second decade of the twentieth century. Thus it happened that the most striking feature of the extra-galactic nebulae was not discovered until about thirty years ago, as a result of the momentous pioneer research of Dr. V. M. Slipher of the Lowell Observatory. In 1912 Slipher first obtained the spectrum of the great nebula in Andromeda and announced that it was moving towards the solar system at the rate of 125 miles per second. By 1917 he had photographed the spectra of fifteen spirals, all but two of which he found receding with velocities averaging about twenty-five times the average velocity of the stars or about 400 miles per second. Although these velocities were enormous, the standard procedure of attributing the observed spectral shifts to the Doppler effect was not immediately called in question. Indeed, no other interpretation appeared to be acceptable. Whatever the ultimate explanation may prove to be, the fact that the spectra of these nebulae are shifted bodily towards the red is indisputable. This extraordinary phenomenon was of crucial significance for the classification of the nebulae, for the gaseous type exhibited no pronounced systematic spectral shift.

The velocities attributed by Slipher to the nearer spirals, although enormous compared with stellar velocities, did not exceed one per cent. of the velocity of light. However, the fact that the vast majority of these nebulae appeared to be receding from our neighbourhood was difficult to reconcile with the idea that their motions were purely random. It was therefore necessary to determine whether this phenomenon applied to more distant nebulae. Unfortunately, most of these are so faint that adequate spectrograms can only be obtained with the aid of the 100-inch reflector at Mount Wilson, and exposures lasting several nights were essential.¹

In 1929 Humason began to photograph the spectra of increasingly fainter nebulae with this great telescope. Within a few years he had amassed an amazing series of photographs indicating a progressive reddening of spectra with increasing faintness and hence with increasing distance. At the limits of observation with his instruments Humason recorded red-shifts corresponding to velocities of recession of nearly 25,000 miles per second or about one-seventh of the velocity of light. As Hubble has reminded us, such velocities are equivalent to circumnavigating the earth in a second or travelling to the moon in ten seconds. Moreover, red-shifts appear to increase still further beyond the limit of the spectrograph.

Thus the existence of a systematic red-shift increasing with increasing faintness has been definitely established. The law correlating shift with distance was first predicted theoretically; it was not until 1929, when Hubble had devised his brightest star criterion of distance, that a sufficient quantity of reliable data could be assembled to confirm that there is a linear relation between distance and shift, known as Hubble's Law. If one nebula is twice as far as another, then its spectrum will be shifted to the red twice as far relatively, and so on.

At first this law was known to be valid only out to distances of about 6 million light-years, but by 1931 it had been ex-

¹ This was a tremendously difficult undertaking. First one point of light among thousands of others all around it had to be kept steadily fixed over the slit of a spectrograph attached to the telescope for, perhaps, eight or ten nights. Then, after all this concentrated effort, the resulting picture might be no more than one-tenth of an inch long and one-thirtieth of an inch wide containing many closely-packed lines!

tended to a distance of the order of 150 million light-years. Further extension of the law is difficult, because of the uncertainties in estimating the distances of the fainter nebulae. In particular, correcting factors in the calculation of distance from apparent brightness will vary according to the interpretation of the red-shift phenomenon. For a nebula receding with a high velocity will appear fainter than a similar nebula which is at rest relative to us: recession will thin out the incoming stream of radiation and hence reduce the apparent brightness by an amount proportional to the red-shift. However, if the systematic red-shifts are interpreted as Doppler shifts, we certainly find that the nearer extra-galactic nebulae, at least, are receding from us in all directions with velocities proportional to their present observed distances, the rate of increase of velocity with distance being about 100 miles per second per million light-years of distance; the farther nebulae may be receding even faster.

On this interpretation we arrive at an extraordinary and unique state of affairs, indicating that the system of extra-galactic nebulae was at one time compressed into a comparatively small region of space in our neighbourhood. *Assuming that the radial velocities have remained sensibly constant*, we calculate that about 1,800 million years ago the nebulae were all clustered together; then 'something happened', and they all began to rush away in all possible directions with all possible speeds. As it is in the highest degree unlikely that our own Galaxy is a peculiar centre of repulsion, we are driven to the conclusion that, on this view, the whole system of nebulae is expanding.

Whatever the ultimate verdict on this conclusion, it is clear that the system of nebulae is markedly different from the Galaxy. Moreover, unlike the Galaxy, this system shows no sign of thinning out at great distances. In each direction we find that increasing numbers of nebulae are visible right up to the very threshold of observation. Possibly we have at last encountered the ultimate background of the universe—the system of extra-galactic nebulae, extending to the horizon of space accessible to our present instruments. If we accept the view that this system is in process of expansion, we are immediately

confronted with the possibility that the universe has existed in its present state only for a period of the order of 2,000 million years.

It is not surprising that this picture of the world has not met with universal acceptance. Attempts have been made repeatedly to find other possible explanations of the red-shift phenomenon. For example, Zwicky suggested that light automatically loses energy in traversing the gravitational field of matter distributed between us and the nebulae, but his calculations have not been generally accepted. Until comparatively recently there seemed to be but two alternative solutions to the problem. These have been well summarized by Hubble as follows: "Red-shifts are produced either in the nebulae, where the light originates, or in the intervening space through which the light travels. If the source is in the nebulae, then red-shifts are probably velocity-shifts and the nebulae are receding. If the source lies in the intervening space, the explanation of red-shifts is unknown, but the nebulae are sensibly stationary."¹¹

Unfortunately, *at present*, the crucial experimental test for deciding between these two possibilities appears to involve us in a vicious circle. For, as already mentioned, a receding nebula should appear fainter than a stationary nebula of the same intrinsic luminosity at the same distance. Hence it is suggested that, if the distance of a nebula is known, its apparent faintness could be used as a criterion of motion. Unfortunately, as Hubble points out, our only information *at present* concerning great distances is derived from the same apparent luminosities which we wish to test. Actually the situation is even more complex than portrayed by Hubble, for he assumes that the average absolute luminosity of the more distant nebulae is the same as that of the nearer. Although the variation in intrinsic brightness of the nearer nebulae is fairly small, being very much less than that of the stars, it is not a necessary consequence of the principle of the uniformity of nature that the average intrinsic brightness of the more distant nebulae *must* also be the same. The light received from the latter probably gives us information about the state of these nebulae hundreds of millions of years ago, and it is possible that the luminosity

¹¹A different possible solution is discussed in the latter part of Chapter VI.

of any given nebula varies appreciably during such vast intervals of time.

We thus see that the distance and motion of the nebulae cannot be measured directly without hypothesis. Indeed, the whole problem bristles with complications, although in time new criteria may help to resolve them. At present, by combining observation and hypothesis we can construct various world-models, and our preference for any one of these will depend on theoretical considerations. Consequently, there is far less agreement about the properties of the system of nebulae than now prevails about the Galaxy.

There is, however, one outstanding feature of the system of nebulae on which there is already fairly general agreement. It arises from counts of nebulae to various limits of faintness. These counts have been made in two different ways. At the Mount Wilson and Lick Observatories, large telescopes were used to study small areas systematically scattered over the sky. These areas were defined by cones with vertices at the observer, and of course yielded information only about areas chosen in the northern sky. Meanwhile, at the Harvard College Observatory and at its southern station, large areas of the sky were studied with the aid of moderate-size telescopes.

Counts of nebulae to various limits of apparent faintness made in these two different ways *at first* appeared to yield discordant results. Shapley, who was responsible for the Harvard programme, claimed that there was pronounced evidence of a general density-gradient; he maintained that, as our telescopes sweep across the sky, they find more nebulae in one region than in another. On the other hand, Hubble and his co-workers, who were responsible for the Mount Wilson programme, claimed that in the three-quarters of the sky accessible to their telescopes there was no evidence of any real density-gradient. Of course there is an apparent gradient, the nebulae appearing to avoid the zone of the Milky Way and to congregate most thickly towards the galactic poles.

Statistical analysis of this apparent gradient shows that, if a correcting factor is introduced to compensate for the light-absorbing effect of the interstellar 'fog' which is spread throughout our Galaxy, then the corrected distribution is

effectively uniform in each direction. However, as statistical analysis is valid only if large numbers of objects are studied, it is necessary to include many faint nebulae. Hubble therefore claims that the mean results of the two methods are substantially the same, indicating that the overall distribution is isotropic, that is, effectively identical in each direction, so that we are at a centre of symmetry. Despite past controversies, we can regard this result as one of the more reliable features of our knowledge of the extra-galactic nebulae. It is to be hoped that future observations with the 200-inch reflector, now being installed on Mount Palomar, will confirm this view.

The fact that we appear to be centrally situated in the field of the nebulae is of vital theoretical significance. Let us compare our apparent position relative to the system of galaxies with our apparent position in our own Galaxy. Prior to Shapley's work on the distribution of globular clusters the sun was thought to be situated at the centre of our Galaxy, but this was ultimately seen to be a wrong conclusion arising from inadequate knowledge. The Milky Way does not extend in the same fashion in all directions, but is bun-shaped and has an ultimate boundary. Consequently, only a comparatively small percentage of stars can be situated at or near the centre of this system. It is improbable that our sun should happen to be one of these, and in fact we now realize that our sun is not central.

On the other hand, the system of nebulae appears to be spherically symmetrically distributed around us and to have no visible boundary. It is theoretically possible, as we shall see, for an unbounded distribution of matter to have its circumference nowhere and centre everywhere. Thus the apparent spherical symmetry of the system of nebulae and the absence of a visible boundary can be reconciled with our aversion to pre-Copernican anthropocentric conceptions of the universe, provided that the system of nebulae has a multitude of centres. If this is so, it is clear that the system must be unbounded, for a nebula on the boundary could not be centrally situated. But if the system of nebulae is unbounded it must form the ultimate background of the physical universe. We have already referred to this possibility, but the present argument puts it in its proper perspective. Although the evidence is not wholly conclusive,

in the present state of our knowledge this assumption forms the basis of almost all current cosmological theories.

When we pass from the consideration of the symmetry to the problem of the distribution in depth of the system of the nebulae, we find no general agreement among the authorities. We have already remarked that there is no sign of any 'thinning out' of the system as we penetrate into the remotest depths accessible to the largest telescopes. Strictly speaking, however, this statement needs some further explanation and qualification. It is true that so far there is no sign of the system 'petering out' at the limits of our vision, but the problem of deciding whether or not there is any actual 'thinning out' of the system at great distances is somewhat involved.

The crude, i.e. 'uncorrected', observational data enable us to draw up a table of the total numbers of nebulae, seen in a particular square degree of the sky, brighter than certain limiting magnitudes, e.g. 18th magnitude, 19th magnitude and so on. These magnitudes represent different degrees of apparent faintness, and the crux of the problem concerns the precise significance which we must attach to this concept.

If we interpret the degree of apparent faintness of a nebula as being solely due to the distance which light has traversed in passing from that nebula to the earth, then on certain simple assumptions it appears that the more distant nebulae (of faintness below the 18th magnitude) must be regarded as *more thinly scattered* in space than the nearer, which appear to be distributed in roughly uniform fashion with about one nebula per cube of side 2 million light-years.

But this simple interpretation of apparent magnitude completely neglects the effect of the red-shift, which is most marked for the faintest nebulae. According to current ideas,¹ this shift, whatever its cause, must give rise to a certain diminution in the energy of the light concerned and so be responsible for some of the apparent faintness. Thus only part of this faintness is due to distance. Hence, in analysing the observational data, apparent magnitudes are usually first corrected for the energy effect. Hubble claims that, when this correction is made, there is no apparent thinning out of the system at the

¹ For an alternative conception see the final paragraphs of Chapter VI.

limits of our telescopes, as the estimated distances of the faintest nebulae are correspondingly reduced. The system now appears to be sensibly *uniform*; in other words, as we examine wider and wider spheres of the universe the number of nebulae observed is roughly proportional to the volume of our field of vision.

Unfortunately this result depends on a number of arbitrary assumptions, even when we have disposed of all 'local' effects, such as the dimming effect of interstellar light-absorbing material. For, in addition to the hypothesis we adopt regarding the cause of the red-shift, we must also make other assumptions, particularly concerning the average absolute brightness of the more distant nebulae which we observe as they were long ago in *time*, and also some assumption about the geometry of *space*. None of these factors had to be considered in deriving the general spherical symmetry of the system.

As we have already remarked, Hubble has examined the respective consequences of assuming that the red-shift is due either to motion in the line of sight or else to some unknown cause. In either case, he assumes that the distant nebulae are of the same average absolute luminosity as those in our immediate neighbourhood, i.e. that this average absolute brightness is sensibly unaltered over periods of hundreds of millions of years. On this assumption and a simplifying assumption concerning the geometry of space, Hubble finds that, if we assumed that the nebulae were approximately stationary, then the observed law of red-shifts would be sensibly linear; in other words, red-shifts would be a constant multiple of distances, so that each unit of light-path would cause the same amount of reddening. This might possibly be regarded as an effect of 'ageing', although Hubble prefers to describe the cause in this case as unknown.

If, however, the nebulae were receding, then those which are more distant would appear to be fainter than if they were stationary. The greatest red-shifts measured so far are about thirteen per cent. of the normal wave-lengths of the incoming light. Hence, these nebulae on this assumption would appear about thirteen per cent. fainter than if they were stationary. The observed law of red-shifts would no longer be linear, for

the farther nebulae would now be regarded as nearer than before, although their red-shifts would be the same. As the light from these nebulae left them several hundred million years ago, Hubble maintains that on this assumption the rate of expansion of the system would have been greater then than now, and hence in this case the time intervals since the expansion began must be shorter than the 1,800 million years previously estimated.

If the apparent magnitudes of the nebulae are corrected merely for the effect of the red-shift in diminishing the energy of their observed light, we have seen that Hubble claims that the system is uniformly spread out in space. If, however, the nebulae are receding, an additional dimming factor arises, and the corresponding correction for distance when incorporated in the calculations destroys the homogeneity. Instead, the number of nebulae per unit volume of space now appears to *increase* as we recede towards the confines of the visible universe. Rightly or wrongly, Hubble maintains that such a picture would imply that we were in a privileged position in the universe, being in the region least densely populated with nebulae. On these, and other grounds, he is inclined, therefore, to reject the Doppler-effect interpretation of the red-shifts and to regard the nebulae as stationary.

Other authorities, notably McVittie and Heckmann, have severely criticized Hubble's method of analysis, but the two alternative models which he presents may possibly be regarded as forming two extremes between which the real universe may lie or even as two different highly simplified ways of looking at the same system. We shall encounter a similar situation in a different context later: on the one hand we shall have an effectively infinite model, homogeneous and isotropic, in which on the average the nebulae are at relative rest; on the other hand we shall be presented with a system which expanded from an initial state of extreme concentration about 2,000 million years ago and now occupies a sphere of radius of the order of 2,000 million light-years. Recalling Hubble's own words we shall find that "A choice is presented, as once before in the days of Copernicus, between a strangely small, finite universe and a sensibly infinite universe plus a new principle of nature."

However justified the criticism of Hubble's analysis may be, his work demonstrates that pure observation unaccompanied by theoretical interpretation tells us very little. In particular, we recall how the revolution wrought by Copernicus found its inspiration in theory rather than in observation. This does not mean that we should underestimate the value of the data which had been laboriously accumulated by observational astronomers. This enormous volume of research was the raw material with which Copernicus constructed his theory, but mathematical ideas and new theoretical concepts were required to give it shape.

Copernicus was concerned with the solar system, the foreground of the universe. Today we believe we have encountered the ultimate background of the universe, the nebulae. Again we are presented with a complicated tangle of observation and hypothesis. In time more powerful means of observation may assist in resolving some of the current controversies, but a clear understanding of the problem will always depend as much on fruitful theoretical concepts, freely constructed by the intellect, as on 'stubborn and irreducible' observational data.

CHAPTER III

SPACE AND TIME

So far we have considered some of the principal results of recent exploration of the universe with the telescope and spectroscope. In assembling the observational evidence we have found that different pictures emerge depending on our method of interpretation. Concerning the general nature of scientific method there are two conflicting points of view which we will consider in turn.

In the early seventeenth century Francis Bacon maintained that Natural Science consisted solely in the patient accumulation of facts. Later in the same century Isaac Newton defined his method of scientific investigation in the famous epigram, *Hypotheses non fingo*.¹ The extraordinary success of modern Natural Science, the foundations of which were laid in the seventeenth century, has caused widespread acceptance of this empirical philosophy. The ultimate court of appeal is said to be the data of observation, which are considered as absolute and independent of the human mind. In the nineteenth century this positivistic point of view was almost universal among men of science. The reaction against it has set in only since the scope of scientific investigation has been so vastly extended on the scale of the very large and of the very small that the classical methods of interpreting phenomena have been found inadequate.

The alternative point of view was first enunciated by Kant. In a famous passage in the preface to the second edition of *The Critique of Pure Reason* he wrote:

“When Galileo caused balls, the weights of which he had himself previously determined, to roll down an inclined plane, when Torricelli made the air carry a weight which he had calculated beforehand to be equal to that of a definite column of water; or in more recent times, when Stahl changed metal

¹ “I make no hypotheses.”

into lime, and lime back again into metal, by withdrawing something and then restoring it, a light broke upon all students of nature. They learned that reason has insight only into that which it produces after a plan of its own, and that it must not allow itself to be kept, as it were, in nature's leading-strings, but must itself show the way with principles of judgment based upon fixed laws, constraining nature to give answer to questions of reason's own determining. Accidental observations, made in obedience to no previously thought-out plan, can never be made to yield a necessary law, which alone reason is concerned to discover. Reason, holding in one hand its principles, according to which alone concordant appearances can be admitted as equivalent to laws, and in the other hand the experiment which it has devised in conformity with these principles, must approach nature in order to be taught by it. It must not, however, do so in the character of a pupil who listens to everything that the teacher chooses to say, but of an appointed judge who compels the witnesses to answer questions which he himself has formulated. Even physics, therefore, owes the beneficent revolution in its point of view entirely to the happy thought, that while reason must seek in nature, not fictitiously ascribe to it, whatever as not being knowable through reason's own resources has to be learnt, if learnt at all, only from nature, it must adopt as its guide, in so seeking, that which it has itself put into nature. It is thus that the study of nature has entered on the secure path of a science, after having for so many centuries been nothing but a process of merely random groping."

This second point of view throws light on the present controversies about the observations of the extra-galactic nebulae. These have not arisen solely because of the inadequacy of the observational data, but also because the nebulae appear to constitute the ultimate background of the physical universe, about which we are obliged to make preliminary assumptions. A satisfactory theory of the structure of the universe must depend on an adequate analysis of these pre-suppositions, in particular the ideas of space and time.

The origin of these ideas is lost in the mists of prehistory. The study of contemporary primitive races, however, suggests

that the earliest ideas were probably rather complicated. It seems that for primitive people time consisted of disconnected fragments; there was no concept of time as a whole, only of convenient fractions of time, e.g. so many moons. The earliest civilization arose as a result of the invention of agriculture, which must have focussed attention on periodical phenomena in nature; indeed, the oldest civilizations are remarkable for their highly elaborate calendars. Nevertheless, the methods of measuring short intervals of time were extremely crude. It will be recalled that as late as the end of the sixteenth century Galileo used his pulse-beat to discover the regularity of the swinging pendulum. In the earliest methods of time reckoning, there were no hours or shorter intervals. Also there was no precise idea of an instant; for example, the shortest interval of time indicated by the Egyptians was depicted in hieroglyphics simply by the upraised head of a hippopotamus with a line cutting it, presumably signifying the period during which the animal thrust its head out of the water for a quick look round. The idea of an hour was developed as a definite fraction of the period of daylight and varied in length according to the season; the hour of fixed length was introduced only with the advent of the striking clock in the fourteenth century.

Similarly, the earliest ideas of space were also extremely complex. In addition to the all-important distinction between up and down, there appeared to be little similarity between one direction and another. We believe that Geometry, or the science of space, was developed originally by the Egyptians and others for the purposes of surveying. Its subsequent transformation by the Greeks into a general abstract discipline, based on axioms and definitions, was one of the most remarkable revolutions in the history of thought. This free creation of the intellect depended on a drastic simplification of the concept of space, thereby artificially attributing to it ideas of homogeneity and immutability. There appears, however, to have been a significant difference between the space of the Greek geometers and the space of the Greek philosophers of Nature. The latter was more in accord with primitive ideas. It was dominated by the idea of 'place': every natural object has its natural position in the universe, towards which it tends to move unless constrained.

Thus in natural space there was absolute position and absolute motion.

The idea of the universe as a whole appears to have been another original creation of the Greek intellect. The Babylonians believed in an 'inevitable necessity' dominating all things, but they do not seem to have had any clear idea of the unity of the universe. As far as we know, the first cosmologist was Thales of Miletus, who lived in the sixth century B.C. He maintained that 'everything is water'. Thus the idea of the unity of the universe was first expressed in terms of an underlying invariant substance. This tendency to abstract something fixed from the eternal flux of phenomena persisted in ancient thought and ultimately resulted in the construction of geometrical models of the universe, in which every object had its allotted position, the whole system being eternal.

The idea of time played a comparatively minor role in Greek thought. It was not refined and developed to the same extent as the idea of space. This comparative neglect of time has had a tremendous subsequent influence and has been responsible for the basically geometrical character of most systems of natural philosophy. The reasons for this neglect were manifold; in particular, it appeared that time was not easily amenable to logical analysis. For example, Zeno claimed to show by a set of ingenious paradoxes that the ideas of time and motion were self-contradictory. In fact, these paradoxes do not disprove the existence of motion, but instead expose certain difficulties associated with the abstract concept of an instant of time. This is seen, for example, most clearly in the paradox of the arrow. Zeno argued that an arrow is at rest during its flight since at each instant it occupies a space equal to itself; but what always occupies a space equal to itself is not in motion and is therefore at rest.

Zeno devised his paradoxes with the object of confuting the followers of Pythagoras. Pythagoras and his school maintained that the ultimate realities in nature are numbers. This doctrine was put forward by Pythagoras after the discovery of the numerical ratio between the length of a string and the pitch of its note. The Pythagoreans were extremely interested in the periodicity of the universe, and one of them, Archytas of

Tarentum, who was a contemporary of Plato, is noteworthy for his definition of time: "Time is the number of a certain movement and in its widest sense the interval of the natural order of the universe." Space was regarded by the Pythagoreans as an assemblage of points, although points were considered to be of finite size and not dimensionless. Every object was thought to be composed of a certain number of points arranged in a definite order.

By exposing the logical difficulties inherent in the Pythagorean ideas, Zeno gave strong support to the rival theory of Parmenides, who maintained that the world was a continuous indivisible sphere always identical with itself. Despite the obvious incompatibility of this concept with our most elementary sense data, this theory has had a profound influence on the development of human thought, causing attention to be directed to those features of the world which are independent of the flow of time, and emphasizing the unity of the universe.

For the Greeks time appears to have been primarily an objective phenomenon characterizing the appearance of the external world. Questions concerning subjective or psychological time were first raised by St. Augustine, whose quaint but illuminating discussion foreshadowed the modern point of view. His famous query, "*Quid est tempus? Si nemo a me quaerat, scio, si quaerenti explicare velim, nescio!*"¹ signified that he was unable to construct a theory of his intuition of time and could not relate it to the objective time-order created by God. The problem concerned him profoundly. "My soul is on fire," he cried, "to know this most intricate enigma."

During the Middle Ages time was regarded teleologically. Instead of picturing it as flowing from past to future, mediaeval philosophers depicted it as the flow of the future (potentiality) into the present (actuality). The whole theory was entirely qualitative.

Modern physics originated with Galileo (1564-1642), who laid the foundations of Dynamics, or the science of bodies in motion. He introduced a new concept of time and its measurement by relating the idea of time to the idea of space. He

¹ "What then is time? If no-one asks me, I know; if I wish to explain it to one that asketh, I know not!"

endeavoured to extend the methods of Euclidean geometry to problems involving the passage of time, which he regarded as measurable by a point moving along a straight line. As points are dimensionless, this idea of time depended on the notion of the durationless instant which the Greeks had rejected on logical grounds. Excessive attention to logical perfection, however, had been responsible for their neglect of problems concerning motion.

Galileo had raised the concepts of space and time to the status of fundamental categories by directing attention to the mathematical description of motion. The mediaeval qualitative method had made these concepts relatively unimportant, but in the new mathematical philosophy the external world became a world of bodies moving in space and time. In the *Timaeus* Plato had expounded a theory that outside the universe, which he regarded as bounded and spherical, there was an infinite empty space. The ideas of Plato were much discussed in the middle of the seventeenth century by the Cambridge Platonists, and Newton's views were greatly influenced thereby. He regarded space as the 'sensorium of God' and hence endowed it with objective existence, although he confessed that it could not be observed. Similarly, he believed that time had an objective existence independent of the particular processes which can be used for measuring it. In the *Principia* he expressed his ideas thus:

"I. Absolute, true and mathematical time, of itself, and from its own nature, flows equably without regard to anything external, and by another name is called duration: relative, apparent and common time is some sensible and external (whether accurate or unequal) measure of duration by means of motion, which is commonly used instead of true time; such as an hour, a day, a month or a year.

"II. Absolute space, in its own nature, without regard to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces which our senses determine by its position to bodies, and which is vulgarly taken for immovable space; such is the dimension of a subterraneous, an aerial, or celestial space, determined by its position in respect of the earth. Absolute and

relative space are the same in figure and magnitude; but they do not remain always numerically the same. For if the earth, for instance, moves, a space of our air, which relatively and in respect of the earth always remains the same, will at one time be one part of the absolute space into which the air passes; at another time it will be another part of the same, and so, absolutely understood, it will be perpetually mutable.

"III. Place is a part of space which a body takes up and is, according to the space, either absolute or relative. . . .

"IV. Absolute motion is the translation of a body from one place into another; and relative motion the translation from one relative place into another. Thus in a ship under sail, the relative place of a body is that part of the ship which the body possesses; or that part of its cavity which the body fills, and which therefore moves together with the ship; and relative rest is the continuance of the body in the same part of the ship, or its cavity. But real, absolute rest is the continuance of the body in the same part of that immovable space in which the ship itself, its cavity and all that it contains, is moved. Wherefore, if the earth is really at rest, the body, which relatively rests in the ship, will really and absolutely move with the same velocity which the ship has on the earth. But if the earth also moves, the true and absolute motion of the body will arise, partly from the true motion of the earth in immovable space; partly from the relative motion of the ship on the earth; and if the body moves also relatively in the ship, its true motion will arise, partly from the true motion of the earth in immovable space, and partly from the relative motions as well of the ship on the earth as of the body on the ship; and from these relative motions will arise the relative motion of the body on the earth. . . .

"Absolute time, in astronomy, is distinguished from relative by the equation or correction of the vulgar time. For the natural days are truly unequal, though they are commonly considered as equal, and used for a measure of time; astronomers correct this inequality for their more accurate deducing of the celestial motions. It may be that there is no such thing as an equable motion, whereby time may be accurately measured. All motions may be accelerated and retarded; but the true or

equable progress of absolute time is liable to no change. The duration or perseverance of the existence of things remains the same, whether the motions are swift or slow, or none at all: and therefore it ought to be distinguished from what are only sensible measures thereof; and out of which we collect it, by means of the astronomical equation. The necessity of which equation, for determining the times of a phenomenon, is evinced as well from the experiments of the pendulum clock, as by eclipses of the satellites of Jupiter.

"As the order of the parts of time is immutable, so also is the order of the parts of space. Suppose these parts to be moved out of their places, and they will be moved (if the expression may be allowed) out of themselves. For times and spaces are, as it were, the places as well of themselves as of all other things. All things are placed in time as to order of succession; and in space as to order of situation. It is from their essence or nature that they are places; and that the primary places of things should be movable is absurd. These are, therefore, the absolute places, and translations out of those places are the only absolute motions.

"But because the parts of space cannot be seen, or distinguished from one another by our senses, therefore in their stead we use sensible measures of them. For from the positions and distances of things from any body considered as immovable, we define all places: and then with respect to such places, we estimate all motions, considering bodies as transferred from some of these places into others. And so instead of absolute places and motions we use relative ones; and that without any inconvenience in common affairs; but in philosophical disquisitions, we ought to abstract from our senses and consider things themselves, distinct from what are only sensible measures of them. For it may be that there is no body really at rest, to which the places and motions of others may be referred."

As Mach pointed out over half a century ago, in these reflections Newton acted contrary to his expressed intention to reject hypotheses and only to investigate actual facts. Absolute time, absolute space and absolute motion are pure mental constructs that cannot be produced in experience. All our experimental knowledge concerns the relative positions and

motions of bodies. However, Newton believed that he had observed absolute motion, and absolute motion implies absolute space and time:

"The effects which distinguish absolute from relative motion are centrifugal forces, or those forces in circular motion which produce a tendency of recession from the axis. For in a circular motion which is purely relative no such forces exist, but in a true and absolute circular motion they do exist, and are greater or less according to the quantity of the absolute motion.

"For instance. If a bucket, suspended by a long cord, is so often turned about that finally the cord is strongly twisted, then is filled with water, and held at rest together with the water; and afterwards, by the action of a second force, it is suddenly set whirling about the contrary way, and continues, while the cord is untwisting itself, for some time in this motion; the surface of the water will at first be level, just as it was before the vessel began to move; but, subsequently, the vessel, by gradually communicating its motion to the water, will make it begin sensibly to rotate, and the water will recede little by little from the middle and rise up at the sides of the vessel, its surface assuming a concave form (as I have experienced) and the swifter the motion becomes, the higher will the water rise, till at last, performing its revolutions in the same times as the vessel, it becomes relatively at rest to it. . . . At first, when the *relative* motion of the water was greatest, that motion produced no tendency whatever of recession from the axis; the water made no endeavour to move towards the circumference, by rising at the sides of the vessel, but remained level, and for that reason its true circular motion had not yet begun. But afterwards, when the relative motion of the water had decreased, the rising of the water at the sides of the vessel indicated an endeavour to recede from the axis; and this endeavour revealed the real motion of the water, continually increasing, until it had reached its greatest point, when *relatively* the water was at rest in the vessel. . . ."

The results of this famous experiment can be summarized thus. When the pail begins to spin there is relative motion between the water and the pail. Gradually the water takes up the motion. Then the pail is suddenly stopped. If all motion

is purely relative, there should be no physical difference between the configuration of the water at the instant when the pail begins to move and at the instant when its motion is arrested. In fact, the surface of the water is level at the former instant and concave at the latter. Newton concluded that rotation must be absolute.

The crucial point in this argument is the implicit assumption that the experiment would yield the same result if it were performed in empty space. In the experiment carried out by Newton the pail was at first rotating and then at rest *relative to the earth*. This criticism was made by Berkeley in his essay *De Motu*. He showed that the motion of the pail only appears to be circular, but that in fact it is very far from being circular because it is necessary to take into account the rotation of the earth about its axis, the revolution of the earth about the sun and so on. Berkeley concluded that it is not necessary to introduce the idea of absolute rotation and that all such motion can ultimately be related to the system of stars as frame of reference.

"Absolute space," said Berkeley, "is infinite, immobile, indivisible, not perceptible by the senses, unrelated to anything, without distinction between its parts. Thus its attributes are negative, it is mere nothing. . . . If every place is relative then every motion is relative, and as motion cannot be understood without a determination of its direction which in its turn cannot be understood except in its relation to our or some other body . . . therefore, if we suppose that everything is annihilated except one globe, it would be impossible to imagine any movement of that globe. Let us imagine two globes and that besides them nothing else material exists, then the motion in a circle of these two globes round their common centre cannot be imagined. But suppose that the heaven of fixed stars was suddenly created and we shall be in a position to imagine the motion of the globes by their relative position to the different parts of the heaven."

Nevertheless, the concept of absolute space as a 'substance' existing independently of matter was retained in classical physics, although modified by certain relativistic concepts. For example Newton, following Galileo, regarded uniform motion

in a straight line as essentially relative. In controverting the arguments brought against Copernicus by those who refused to believe that the earth could move, Galileo had pointed out that if a ship is moving evenly with any constant velocity it is impossible to judge from the behaviour of objects on board whether the ship is moving or not. Galileo's idea was expressed by Newton in the form: "The motions of bodies enclosed in a given space are the same relatively to each other whether that space is at rest or moving uniformly in a straight line without circular motion." Thus in Newtonian empty space rotation is absolute, but motion in a straight line with uniform speed is relative. In other words, experiments can be performed to detect rotation, but no corresponding experimental test can be devised to indicate absolute uniform motion in a straight line.

A radically different theory of space and time was put forward by the great German mathematician and philosopher Leibniz (1646-1715). He rejected the idea that space and time exist independently and in their own right. He believed that laws of nature are laws of thought, but he drew a distinction between truths of pure reason, which are 'necessary' because their opposites are self-contradictory, and truths of fact which are 'contingent' because only a 'sufficient' reason can be given why they should be so and not otherwise. He regarded propositions dealing with physical existence as being of this nature. Only a sufficient reason can be assigned to them because their opposites are not self-contradictory. By applying this principle, he came to the conclusion that space and time are not 'things' but orders or methods of arranging objects and events. In a letter to Clarke, the English defender of the Newtonian principles, Leibniz wrote:

"I say then that if space were an absolute being, there would happen something for which it would be impossible that there should be a sufficient reason. . . . Space is something absolutely uniform, and without the things situated in it one point of space does not differ in any respect from another point of space. Now from this it follows that if we suppose that space is something in itself, other than the order of bodies among themselves, it is impossible that there should be a

reason why God, preserving the same positions for bodies among themselves, should have arranged bodies in space thus and not otherwise, and why everything was not put the other way round (for instance) by changing east and west. But if space is nothing other than this order or relation, and is nothing whatever without bodies but the possibility of placing them in it, these two conditions, the one as things are, the other supposed the other way round, would not differ from one another: their difference exists only in our chimerical supposition of the reality of space itself. . . .

"The same is true of time. Suppose someone asks why God did not create everything a year sooner; and that the same person wants to infer from that that God did something for which He cannot possibly have had a reason why He did it thus rather than otherwise, we should reply that his inference would be true if time were something apart from temporal things, for it would be impossible that there should be reasons why things should have been applied to certain instants rather than to others, when their succession remained the same. But this itself proves that instants apart from things are nothing, and that they only consist in the successive order of things; and if this remains the same, the one of the two states (for instance that in which the creation was imagined to have occurred a year earlier) would be nowise different, and could not be distinguished from the other which now exists."

These revolutionary concepts of space and time played little part in the development of physics for over two hundred years, and a further new idea was required before their physical, as distinct from their philosophical, significance could be appreciated: the idea of 'the observer'. This idea originated in the famous observations of Roemer in 1676, establishing that light takes time to travel. Previously, Galileo had attempted to discover if light had a finite velocity, but his experiments were too crude, consisting of simple signalling by means of a lantern on a hilltop. Roemer's method was astronomical; he attempted to deduce the velocity of light from the observation of the eclipses of Jupiter's moons. These eclipses occur frequently, the two nearest moons to Jupiter being eclipsed during each revolution, the periods of revolution being $42\frac{1}{2}$ hours and $85\frac{1}{2}$

hours respectively. According as Jupiter and the earth are on the same or opposite side of the sun respectively, so the distance of Jupiter from the earth varies by an amount which is approximately equal to the diameter of the earth's orbit. Roemer discovered that the eclipses of Jupiter's moons were observed earlier when Jupiter was near the earth and later when Jupiter was farther away. He suggested that this time-difference in the eclipses was equal to the time light takes to traverse the earth's orbit. In this way he calculated the velocity of light to be 192,000 miles per second, compared with the value of 186,000 miles per second now generally accepted.

Eddington has remarked that time as we now understand it was discovered by Roemer. For his discovery revealed that the world-as-seen at a given epoch depends on the location of the observer, and that there is no unique correlation between the succession of instants in our consciousness and the succession of events which we witness in the external world. The apparent order of events depends on who is observing them. For example, today we see two phenomena in the heavens: a prominence appears on the sun and a new star flares up in one of the extra-galactic nebulae; but the former event happened less than ten minutes ago, while millions of years have elapsed since the latter. It is not easy to reconcile with this situation the concept of an objective time-order existing independently of the observer.

A new theory uniting the psychological and physical concepts of space and time was suggested in the latter half of the eighteenth century by Immanuel Kant. He maintained that space and time are forms of thought; they are ideas existing in our minds prior to any observation of phenomena, moulds into which we pour the data of physical experience. Thus Kant, like Berkeley, regarded space and time as subjective, i.e. dependent on the observer, but he did not believe that these concepts were derived solely from experience. Similarly, he agreed with Leibniz in regarding space and time as orders of phenomena, although he did not consider these orders to be objective, i.e. independent of the observer. He also retained Newton's idea of absolute space, "not as a conception of a real object, but as a mere idea which is to serve as a rule for con-

sidering all motion therein as merely relative". Although Kant's views had little direct influence on the development of classical physics, they had a profound effect on philosophy in general, for they focussed attention, in a fundamental manner, on the role played by the observer or spectator of natural phenomena. This orientation of thought was of more lasting significance than the precise details of his philosophy.

Kant maintained that, as our idea of space is a necessity of thought, geometry or the science of space is a direct consequence of our method of reasoning. When Kant wrote, geometry was still the unique system expounded by Euclid, but in the early nineteenth century it was discovered that there are other systems of geometry which are equally valid logically. One of the basic assumptions of Euclidean geometry is the famous axiom of parallels: through any given point in space one and only one straight line can be drawn parallel to a given straight line. It had long been felt that this axiom was artificial, and numerous attempts had been made to deduce it from the other more cogent axioms of geometry. As failure followed failure, it occurred to certain geometers, notably Lobatchewski in Russia, Gauss in Germany and Bolyai in Hungary, that it might be possible to construct valid systems of geometry to which the parallel postulate did not apply. Replacing Euclid's axiom by the postulate that an infinite number of parallels to a given line can be drawn through a given point, Lobatchewski discovered a system of geometry which is quite distinct from that of Euclid. Other systems were discovered later.

During the remainder of the nineteenth century there was a fundamental cleavage of opinion between the philosophers and the mathematicians concerning the nature of space, although the old idea that physical space exists in its own right still persisted. The philosophers regarded it as Euclidean and believed that its geometry could be established by pure thought, while the mathematicians maintained that the geometry of physical space could only be determined empirically, for example by astronomical observation.

At the close of the century a new way to reconcile the two points of view was suggested by the great French mathematician Henri Poincaré. He believed that there is no unique

form of geometry which we are obliged to assign to the external world. For example, it had been suggested that the geometry of physical space could be determined, in principle, by measuring the sides and angles of triangles formed by various stars. Poincaré pointed out that such measurements would tell us nothing about space. Instead, they would provide us with information about the light rays travelling between the stars, but we would not be compelled to regard these rays as geometrical straight lines. We could map these rays in any space we cared to choose, although in some spaces we should find simpler geometrical relations associated with these rays than in others. From the various logically possible geometries, we could choose the one which was most convenient for our purpose.

This point has been illustrated by the following picturesque analogy due to the French mathematical philosopher Nicod. He points out that the geometrical pattern of natural events can be read in different ways like a well-known type of picture puzzle. "As children," he says, "we have all seen those pictures which represent things that we cannot distinguish at the first glance; where it is a matter of discerning a giraffe or a lion in the lines of a landscape deserted when first scanned. When we have 'discovered' the picture hidden in them, we have seen nothing new. The contour of this little mountain is now the mane of a lion, and the knot in this tree trunk is its eye. We had read in this network of lines a certain structure, the landscape, and now we have just read a second structure, the lion." Similarly, we can associate various geometries with the external world, depending on how we choose to identify particular natural phenomena with particular geometrical concepts, e.g., rays of light with straight lines or other curves.

In astronomy the distances and the times of events are determined by the properties of light. As we have seen, the fact that light travels at a finite speed ultimately undermined the age-old belief that there is a unique time-order of events in the universe. Instead, we now believe that this time-order depends on the location of the observer and varies from one observer to another. Similarly, the observer can choose the space in which he maps the geometrical relations between physical phenomena. Thus the concepts of space and time,

which were formerly assumed to be fundamental and unique aspects of the universe, are now regarded as dependent on the concept of the observer and the properties of light. That this new point of view must result in far-reaching modifications of classical physics was first realized by Einstein in 1905 when he formulated his famous Special Theory of Relativity.

CHAPTER IV

RELATIVITY

WE have already mentioned that mechanical forces produce the same effects on bodies in uniform motion as on bodies at rest relative to the earth. For example, any mechanical experiment performed on board a ship sailing steadily in a straight line will yield the same result as a similar experiment carried out on land. Although the earth rotates around the sun and the sun moves with respect to the stars, the motion of the earth with respect to the basic frame of reference¹ of the universe can be regarded as sensibly uniform during the small interval of time covered by a laboratory experiment. Hence the laws of mechanics must be the same, or very nearly the same, in all frames of reference which are in uniform relative motion; consequently it is not possible to devise a mechanical experiment which would permit us to measure the absolute velocity of the earth.

The great advances in optics made during the last century suggested that it might be possible to measure the absolute velocity of the earth by an optical experiment. This was first attempted in 1887 by Michelson and Morley. Their experiment ultimately caused a revolution in our ideas of space and time. The Newtonian concept of an absolute empty space had been severely criticized by philosophers, but these criticisms had had no influence on physicists. The systematic investigation of optical and electrical phenomena in the nineteenth century, however, led to the hypothesis that space as whole was pervaded by the ether, a medium capable of transmitting vibrations and disturbances. In the course of time some extremely curious properties were attributed to this medium. Nevertheless, it appeared to provide a fundamental frame of reference. Michelson and Morley, therefore, attempted to measure the velocity of the earth with respect to the ether.

¹ In Newtonian theory a frame of reference is a system of space and time measurement.

The theory of their experiment was simple. Assuming that light was due to vibrations set up in the ether, they believed that its velocity must be independent of the motion of the source from which it was emitted, depending only on the properties of the ether itself. A similar effect occurs when a stone is thrown into a pond. The motion of the waves set up is independent of the speed of the stone, and is determined solely by the characteristics of the pond.

In the experiment of Michelson and Morley, a source of light was placed at the junction of two equal arms arranged at right angles.¹ Mirrors were fixed at the ends of these arms so that the light could be reflected back to its source. By means of a highly accurate instrument, known as an interferometer, it was possible to detect minute differences in the times of return of these reflected rays emitted simultaneously from the source. One arm of the apparatus was placed in the direction of the earth's rotation in its orbit and the other at right angles. If the earth were moving through ether, the two rays should return to the source at different times. For the to and fro velocities of light relative to the arm which was pointing in the direction of the earth's motion should differ from the velocity relative to the arm which was pointing in the transverse direction. The velocity of the earth in its orbit relative to the sun is about one ten-thousandth of the velocity of light, and it was calculated that the times taken to traverse the two arms should differ by at least two parts in a hundred million. The apparatus was so accurate that far smaller differences could have been detected. However, when the experiment was first performed the two rays returned simultaneously so that the velocity of the earth was given as zero. As this unexpected result might have been due to the earth happening to be at rest relative to the ether at that epoch, the experiment was repeated six months later when the earth's motion in its orbit was completely reversed in direction, but again the same result was obtained.

Numerous attempts were made to explain this failure to measure the earth's motion through the ether. Michelson himself suggested that the earth dragged along the ether in its

¹ In the actual experiment, the arrangement of the apparatus was somewhat more complex than here described.

neighbourhood so that the velocity of the earth with respect to the ether would appear to be zero. Abandoning the ether theory, Ritz showed in 1908 that the experiment could be explained if it were assumed that the velocity of light depended on the velocity of its source. But this hypothesis, in turn, was contradicted by other facts; in particular, the Dutch astronomer de Sitter showed conclusively five years later that the relative motion of double stars failed to produce any differential effect on the velocity of light.

A more revolutionary hypothesis was suggested in 1892 by the Irish mathematical physicist Fitzgerald. He suggested that every body moving relative to the ether automatically contracts by a definite fraction depending on its velocity. Of course, for velocities which are small compared with that of light this fraction must be minute. Thus the null result of the Michelson-Morley experiment was accounted for by assuming that the difference in velocity of the two light rays was compensated by a difference in effective length of the two supposedly equal arms. This explanation, of course, could not be verified by direct measurement because every measuring rod would itself be shortened in the same way.

The Dutch physicist Lorentz showed that the Fitzgerald hypothesis also accounted for the null results of other experiments designed to measure the earth's motion relative to the ether. We now realize that all these experiments indicate that the laws of optical and electromagnetic phenomena must, like the laws of mechanics, be the same for all systems in uniform relative motion. This conclusion was first reached by Einstein in 1905, who suggested that the simplest explanation of the Michelson-Morley experiment was the hypothesis that the velocity of light is the same for all observers in uniform relative motion. This explanation was revolutionary because it was incompatible with the classical ideas of space and time. According to these, two observers in uniform relative motion cannot assign the same finite velocity to any moving object, for the velocities which they assign respectively must differ by their own velocity relative to each other. Hence, on the classical concept, if light moves with a finite velocity its value cannot be the same for all observers in uniform relative motion.

As Einstein's explanation of the Michelson-Morley experiment conflicted with the classical ideas of space and time, he began with a penetrating criticism of these concepts. In particular, he rejected the notion of absolute time as it implies that a meaning can be attached automatically to the simultaneity of two events in different places. He showed conclusively that the idea of two events being simultaneous must depend on some convention of measurement, unless the two events occur at the same place.

For example, let us suppose that an explosion occurs on Mars, which is observed by an astronomer on the earth, who records the instant when he sees the flash. If light travelled instantaneously with an infinite velocity, this instant would coincide with the time of the explosion recorded by a hypothetical observer on Mars. In this way a meaning could be attached automatically to absolute time and the simultaneity of events at different places; indeed, the classical theory is now regarded as the limiting form of Einstein's theory when the velocity of light becomes infinite. But as there is a mass of experimental evidence supporting the view that light takes a finite time to travel a definite distance, the terrestrial observer must correct the time recorded on his watch. This correction for the time taken by light to travel from Mars will depend on assumptions concerning the velocity of light and the measurement of distance. Thus the concept of a world-wide simultaneity ceases to be a primitive idea.

Although Einstein abandoned the classical concept of absolute time, he assumed that each observer possesses a unique standard of time for recording events occurring in his immediate neighbourhood. He also assumed that each observer is provided with a rigid rod for measuring lengths. At first he confined his attention to uniformly moving observers, and proved that, if each observer assigns the same constant velocity to light rays passing between them, precise relations can be deduced correlating the measurements made by any pair. According to either observer, a yard-stick or metre rod used by the other for measuring lengths in the direction of their relative motion will appear to be shorter than one used by himself. Thus Einstein deduced and reinterpreted the Fitzgerald

contraction. Similarly, according to either observer, the standard clock kept by the other will appear to run slow. There will, however, be complete symmetry between the two observers, due to the initial postulate that the laws of nature have the same form for both. In particular, each observer will assign the same relative velocity to the other. This principle of complete symmetry of observers is called the Principle of Relativity, and when they are restricted to uniform relative motion the theory of their interrelations is called the Special Theory of Relativity.

If each observer assigns the velocity v to the other and c is the velocity of light, then according to either observer the length of a standard metre rod carried by the other is shortened by a factor $\sqrt{1 - v^2/c^2}$ corresponding to the Fitzgerald contraction. Similarly, a definite time interval recorded on a clock carried by either observer will appear to the other to be shortened by the same factor. For all velocities encountered in everyday life this factor is so close to unity that it can be neglected. Thus for the velocity of the earth in its orbit this factor differs from unity by about one part in 200 million. Only for velocities which are not small fractions of the velocity of light will the factor differ appreciably from one. We recall that velocities of the order of a seventh of the velocity of light have been attributed to the most distant nebulae whose spectral shifts have been measured. Even for these enormous velocities the contraction amounts only to about two per cent. However, for velocities nearly equal to the velocity of light this factor becomes extremely important and the length of a moving rod appears to tend to zero.

Similarly, a clock carried by an observer moving with a velocity nearly equal to that of light would appear to run very slowly. In fact, the velocity of light plays the role of a limiting velocity, and it is often stated that it would be impossible for an object to have a greater velocity. Of course, this is a consequence of our method of correlating measurements by two observers. For if one observer had a greater relative velocity of recession than that of light it would be impossible for him to receive a light signal from the other. Thus an observer moving in this way could not be brought within the scope of the theory; in

practice, material objects moving with such velocities have not yet been detected.

Since the apparent length of a moving rod depends on who is observing it, we see that the Special Theory of Relativity is incompatible with the classical idea of absolute space. Similarly, the measurement of the time interval between two given events depends on who is measuring it. The time indicated by clocks in the system in which they are at rest is called 'the proper time' of the system. In Lorentz's theory of the electron this proper, or local, time had already appeared as an auxiliary mathematical quantity, but he did not regard it as true physical time. In Einstein's theory, however, there is no means of isolating this time from the infinite number of equivalent local times of different moving systems. Thus Newton's idea of absolute time plays no part, and all statements concerning time have a meaning only when referred to a definite observer.

The Special Theory of Relativity automatically accounted for the null result of the Michelson-Morley experiment. Moreover, it also explained a number of other experimental results which were difficult to incorporate in the framework of classical physics, for example, Fizeau's experiment. It had long been known that light travels through a liquid, or other material medium, with a slower speed than through ether, or empty space. In the middle of the last century the French physicist Fizeau discovered by experiment that if the velocity of light in a particular liquid is u relative to the liquid, and if the liquid is moving with speed v along a tube, then the velocity of the light relative to the tube is less than the sum of u and v , although according to classical ideas it should be equal to this sum.

Einstein showed that this result could be immediately explained by his theory. Whereas in classical physics velocities are compounded by the ordinary law of addition, so that a man walking at three miles per hour along the corridor of a train travelling at sixty miles per hour will appear to a man standing on a station platform to be moving at sixty-three miles per hour; according to the Special Theory of Relativity he will appear to have a slightly lower velocity. Indeed, all velocities are compounded in such a way that the resultant of any number can never exceed the velocity of light in empty space. In par-

ticular, this velocity must be unaltered when compounded with any other velocity, as it is assumed that light has the same speed relative to each uniformly moving observer.

Although in Einstein's theory each observer has his own private space and time, so that different observers assign different lengths to the same object and different measures to the interval of time between two given events, these spaces and times, as we have already remarked, are interrelated by precise equations. The mathematical form of these equations is identical with that of certain equations discovered by Lorentz, and called after him the Lorentz formulae. Einstein, of course, gave them a new and much more powerful physical interpretation. These equations are so important that we must now explain them in detail.

Consider an event, for example the outburst of a nova, or a sudden cataclysmic increase in the brightness of a faint star. Suppose this event is observed from two stars in line with the nova, and suppose further that the two stars are moving uniformly with respect to each other in this line. Let the epoch at which these stars passed by each other be taken as the zero of time measurement, and let an observer *A* on one of the stars estimate the distance and epoch of the nova outburst to be *x* units of length and *t* units of time, respectively. Suppose the other star is moving towards the nova with velocity *v* relative to *A*. Let an observer *B* on this star estimate the distance and epoch of the nova outburst to be *x'* units of length and *t'* units of time, respectively. Then the Lorentz formulae, relating *x'* and *t'* to *x* and *t*, are

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}; \quad t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}.$$

These formulae are, of course, quite general, applying to any event in line with two uniformly moving observers. If we let *c* become infinite then the ratio of *v* to *c* tends to zero and the formulae become

$$x - vt; \quad t' = t.$$

Thus if, and only if, the velocity of light were infinite, would each observer assign the same epoch to the event (so many years from the instant when the observers parted company), in agreement with the Newtonian idea that time is independent of the observer. Similarly, the distances assigned by *A* and *B* to the event would differ only by the distance between *A* and *B* themselves, in accordance with the Newtonian idea of absolute space.

In 1908 the famous mathematician Minkowski made a remarkable discovery concerning the Lorentz formulae. He showed that, although each observer has his own private space and private time, a public concept which is the same for all observers can be formed by combining space and time in a particular way. If we regard an interval of time as a kind of 'distance' in the time dimension, we can convert it into a true distance by multiplying it by the velocity of light, *c*; in other words, with any time interval we can associate a definite spatial interval, namely the distance which light can travel in empty space in that period. If, according to a particular observer, the difference in time between any two events is *T*, this associated spatial interval is *cT*. Then, if *R* is the space-distance between these two events, Minkowski showed that the difference of the squares of *cT* and *R* has the same value for all observers in uniform relative motion. The square root of this quantity is called the space-time interval between the two events. Hence, although time and three-dimensional space depend on the observer, this new concept of space-time is the same for all observers.

Minkowski's discovery provides a vivid interpretation of Einstein's ideas. We now see that if the universe is pictured as a system of events, then the spaces and times of different observers are simply different cross-sections of this system. The space-time interval, which is the same according to all observers, is split up, owing to their different directions and speeds, into different space and time components. Weyl has commented on this idea in the following passage. "The scene of action of reality is not a three-dimensional Euclidean space but rather a four-dimensional world in which space and time are linked together indissolubly. However deep the chasm may be that

separates the intuitive nature of space from that of time in our experience, nothing of this qualitative difference enters into the objective world which physics endeavours to crystallize out of direct experience. It is a four-dimensional continuum, which is neither 'time' nor 'space'. Only the consciousness that passes on in one portion of this world experiences the detached piece which comes to meet it and passes behind it as *history*, that is as a process that is going forward in time and takes place in space."

Einstein's original restriction of the principle of relativity to observers in uniform relative motion was in accordance with the fundamental laws of Newtonian mechanics. Newton's first law of motion¹ stated that "every body continues in its state of rest or uniform motion in a straight line, except in so far as it may be compelled by force to change that state". The relations between observers in accelerated relative motion appeared to be on quite a different footing, for according to classical ideas acceleration could occur only through the action of force. In Newtonian mechanics the assumption that the laws of nature are the same for all frames of reference in uniform relative motion (inertial frames) was reconciled with the idea of absolute space by postulating that in each frame the same length is assigned to the spatial interval between any two given places. Once this postulate had been abandoned, and the notion of absolute space rejected, the restriction of the principle of relativity to frames of reference, or observers, in uniform relative motion seemed artificial. For, if space and time are relative, then presumably acceleration, like velocity, should be relative too. If so, there should be some generalization of Newton's first law to include accelerated motion.

The most familiar example of acceleration in nature occurs when an object falls freely in the earth's field. In attempting to extend the theory of relativity to embrace motions of this type, Einstein considered the case of a man experimenting on gravitation while in a lift. As long as the lift is at rest, the man can determine the strength of the gravitational field on the earth's surface in the usual way. He finds that all objects undergo a downward acceleration of 981 centimetres per second per

¹ The law of inertia.

second until their motion is stopped by collision with other bodies or with the floor of the lift. As this acceleration is found to be the same for all bodies in the lift, the man might conclude that the lift itself is subject to an upward acceleration of 981 centimetres per second per second and that those bodies inside the lift, which, at least during a short time, are not forced to move in the same way, obey Newton's first law of motion and remain behind until the floor of the lift collides with them.

Suppose now that the cable supporting the lift snaps so that the lift begins to fall freely in the earth's gravitational field. Then all the bodies in the lift will undergo the same acceleration as the lift itself and consequently will be unaccelerated relative to each other. Thus to the man in the lift it will appear that the lift itself is no longer in a field of force, but is moving in accordance with Newton's first law and so is either at rest or in uniform relative motion. But a man outside the lift and at rest on the earth's surface will interpret the situation differently. Before the cable snaps he will regard the lift as being at rest, and after it snaps he will regard it as falling with uniform acceleration. Consequently, Einstein asserted that acceleration is not absolute, as it must depend on the choice of observer.

The conclusion that acceleration is relative implies that the concept of force is also relative. To illustrate this point, Einstein considered a closed box, containing an observer and apparatus, placed in a region of space where there is no gravitational field due to outside bodies. To the middle of the lid of the chest a hook is fixed and a rope is attached. A playful spirit begins pulling at the rope with a constant force, so that the box and the observer begin to move with uniformly accelerated motion. How will the observer regard the situation? The acceleration of the box will be transmitted to him by the reaction of the floor. Moreover, if he releases an object which he had been holding, the acceleration of the box will no longer be transmitted to it and it will fall to the floor with a relative acceleration. This relative acceleration will be the same for all objects. Thus it will be natural for the observer to assume that he and the box are in a uniform gravitational field similar to that which conditions existence on the earth. Every experiment which the observer can make will confirm his belief that the

chest is at rest in a gravitational field, whereas to an outside observer it will appear to be moving with uniform acceleration. We see therefore that gravitational force is relative and not absolute, depending on the choice of observer. Consequently, any extension of the principle of relativity to systems in relative acceleration should yield important results concerning gravitation.

In Newton's theory of the world, space was necessarily Euclidean because no other type of geometry had then been invented. Einstein contended that, if all possible hypothetical observers can be introduced into physics, it could no longer be assumed that the geometry of space according to each observer is Euclidean. This is easily seen in the case of two coincident observers *A* and *B* whose frames of reference are in relative rotation. We suppose that a circle is drawn about each observer in the plane perpendicular to the axis of rotation of *B*'s frame. If space is Euclidean according to *A*, then the ratio of the circumference of the circle to its diameter as measured by *A* is the well-known number π . If the circle is also measured by *B*, his rod will give the same length for the radius, but in measuring the circumference *B* will attribute the Fitzgerald contraction to *A*'s rod,¹ and so according to *B* the ratio of the circumference to the diameter will not be equal to π . Hence, if Euclidean geometry applies to the space of *A*, it will not apply to that of *B*. Consequently, in creating his General Theory of Relativity, Einstein needed a more general type of geometry than that of Euclid.

The General Principle of Relativity as formulated by Einstein asserts that the laws of nature must be expressible in a form which is the same for all possible observers moving in any way and hence for the most general transformation of space and time co-ordinates. Minkowski had shown that the fundamental geometry for unaccelerated systems remote from gravitational influence is a four-dimensional space-time. Einstein extended this idea to accelerated systems. He assumed that each portion of space-time has the same magnitude for all observers or moving systems although its space and time com-

¹ Einstein assumed that this will happen if *B*'s frame rotates uniformly with respect to *A*'s.

ponents vary from one observer to another. He postulated that the geometry of space-time must be of the very general kind first studied by the German mathematician Riemann. He also postulated that all gravitational motions take place along the shortest paths in space-time; these are the analogues of straight lines in Euclidean geometry. Space-time as a whole could not be uniform because, for example, the gravitational paths of planets in the field of a double star are not the same as the paths of planets rotating about a single star. In short, Einstein assumed that the geometry of any region of space-time depends on the material objects occupying that region.¹

This idea was not entirely new. In the late nineteenth century the British mathematician Clifford had made a similar suggestion on general philosophical grounds. In his posthumous *Common Sense of the Exact Sciences*, published in 1885, there occurs the prophetic passage: "We may conceive our space to have everywhere a nearly uniform curvature, but that slight variations of the curvature may occur from point to point, and themselves vary with the time. These variations of the curvature with the time may produce effects which we not unnaturally attribute to physical causes independent of the geometry of our space. We might even go so far as to assign to this variation of the curvature of space 'what really happens in that phenomenon which we term the motion of matter'."

In developing his geometrical theory of gravitation, Einstein looked for a particular restriction on the general Riemannian geometry of space-time which, if true for any one observer, would automatically be true for all. If this restriction determined a type of Riemannian geometry in which the shortest paths could be identified with motions in gravitational fields, it could be regarded as a reformulation of the law of gravitation. Einstein succeeded in discovering a suitable condition of this type.

The simplest method of testing this new law of gravitation was to consider the gravitational field surrounding a single particle, as the space surrounding such a particle must be

¹ It can be shown that the shortest path 'postulate' is, in fact, a logical consequence of the general equations, due to Einstein, relating the distribution of matter and the structure of space-time.

symmetrical. If we seek the shortest paths in the space-time associated with this particle, we find that they correspond very nearly to motion in an ellipse, but the ellipse itself rotates about the centre of gravitational attraction. According to Newtonian theory, if we correct the observed motions of the planets for their perturbations upon each other, we ought to obtain a stationary ellipse for the orbit of each planet with respect to the sun. For most of the planets the Einstein deviation from elliptical orbits is extremely small, and only for one planet is the deviation sufficiently large to be regarded as a serious discrepancy with the calculations of Newtonian theory.

The orbits of the planets are in the main very nearly circular, so that it is extremely difficult to locate accurately a definite point, for example the perihelion or point nearest to the sun. However, in the case of Mercury this can be attempted, and about a hundred years ago the great French astronomer Leverrier, co-discoverer with Adams of the planet Neptune, claimed that there was an unexplained observed motion in the perihelion of Mercury amounting to 43 seconds of arc per century. This would correspond to a complete rotation of the whole orbit about the sun in rather more than a million years. To explain this discrepancy between theory and observation without rejecting the Newtonian theory, Leverrier suggested that the orbit might be perturbed by another planet, Vulcan, nearer to the sun than Mercury and hence extremely difficult to observe. Such a planet has never yet been discovered. However, Einstein showed that this hypothesis was unnecessary, as his new law of gravitation implied that the elliptical orbit of Mercury should rotate about the sun at the rate of 42 seconds of arc per century. This explanation of the hitherto inexplicable residuum of the motion of Mercury has been generally regarded as a remarkable confirmation of Einstein's theory. Nevertheless, some authorities maintain that this agreement between observation and theory is partly fortuitous, as the probable error in the observed discrepancy, when corrected for all other possible effects, is an appreciable fraction of 43 seconds.

Einstein suggested two other critical tests of the new theory. One of these concerned the spectra of stars. Since the properties of space and time depend on the properties of matter,

it follows that, in a more intense gravitational field than that of the earth, time should run at a different rate. Thus, according to the General Theory of Relativity, a natural clock on the sun, for example a vibrating atom, should run more slowly than the corresponding clock on the earth. This means that it sends out vibrations at a slower rate. Now the speed at which radiation, for example light, travels is measured by the product of the wave-length and the number of waves emitted in unit time, which is usually taken to be a second. If the frequency or number of waves emitted per second is reduced, then the wave-length must be correspondingly increased, and hence there must be a displacement of the spectrum.

We have already mentioned that, according to the theory of Doppler, the radiation emitted by a body automatically exhibits a spectral shift if the body is moving away from the observer. This phenomenon can be immediately accounted for on the Special Theory of Relativity, since the apparent rate of any clock moving uniformly away from the observer is slowed down, as we have already remarked. The General Theory of Relativity asserts that a strong gravitational field produces a similar effect to recession. In the case of the sun, the predicted displacement is very small, the wave-length of a typical line in the solar spectrum being increased, compared with the wave-length of the corresponding line in the laboratory, by only two parts in a million. This displacement appears to have been confirmed experimentally, but the prediction is far more easily tested in the case of a much denser star. Sirius has an extremely dense companion star, an average cubic inch of which contains over a ton of matter, so that its average density is more than eighty thousand times that of water. For this star the predicted gravitational displacement of the spectrum is about thirty times as great as for the sun; this result was confirmed experimentally by the American astronomer Adams in 1924.

The third test suggested by Einstein concerned the passage of light through strong gravitational fields, for the new theory indicated that light is deflected by such fields. To appreciate this test we must introduce some additional concepts. According to Newton's second law of motion, the deviation from

uniform motion, i.e. the acceleration, of a body depends on the force acting on it, and also on an inherent property of the body which is called its inertial mass. In Newtonian mechanics it was assumed that this quantity is fixed, irrespective of the conditions to which the body is subjected, in particular its motion. This inertial mass, which Newton called the 'quantity of matter' in a body, can be regarded as the measure of the resistance of the body to any force acting upon it.

In developing his theory of gravitation, Newton assigned to each material body another property which is called its gravitational mass. Gravitational mass determines the force exerted by the body on other bodies, and so its function appears to be quite distinct from that of inertial mass. Nevertheless, the two are found to be identical in magnitude. Newton made experiments to verify this remarkable equality by swinging a pendulum with a bob which could be made with different materials. The period of swing depended on the ratio of the inertial and gravitational masses of the pendulum, but in all cases it was found to be the same, thus confirming the equality of the two masses. In 1890 Eötvös made a much more refined test with the aid of a highly sensitive instrument called a torsion balance. Repeated experiments showed that inertial and gravitational mass were equal to within one part in 100 million. Einstein suggested that this was because inertia and gravitation are identical. As we have seen in the case of the man in the lift, the same phenomena can be interpreted equally well in terms of either concept.

According to the Special Theory of Relativity, the velocity of a moving body is always less than the velocity of light. Since the energy of motion of a body depends on its inertial mass and its velocity, it follows that, if the energy of a body is increased indefinitely by the continued application of a force, the inertial mass of the body must be increased too; for, if not, the velocity would ultimately increase indefinitely and exceed the velocity of light. Einstein found that, corresponding to any increase in the energy content of a body, there is an equivalent increase in its inertial mass. Mass and energy thus appeared to be different names for the same thing, the energy associated with a mass M being Mc^2 , where c is the velocity of light; and

the mass M of a body moving with velocity v he found to be given by the formula

$$M = \frac{m}{\sqrt{1 - v^2/c^2}}$$

where m is the rest-mass or mass of the body as measured by an observer with respect to whom it is at rest. This unification of the concepts of mass and energy is the most important consequence of the Special Theory of Relativity. It implies that matter can be regarded as highly concentrated energy, a view which has received remarkable confirmation in recent work on nuclear fission. The enormous release of energy in such processes is due to the conversion of a small quantity of mass into its equivalent amount of liberated energy.

Since the time of Clerk Maxwell light has been regarded as a form of electromagnetic energy. Hence it should possess inertial mass, as is also indicated by the existence of radiation pressure; for it has been shown experimentally that a beam of light exerts pressure on any object on which it falls, and so the object must exert some force on the light in order to stop it. Indeed, radiation pressure is believed to play a preponderant role in determining the sizes of the larger and more diffuse stars, balancing the weight of the peripheral gases. The fact that the tails of comets passing near the sun are always directed away from the sun is also believed to be due to the radiation pressure of sunlight on the minute particles of which these tails are thought to be composed.

Thus there is abundant evidence that light has inertial properties. Hence, if inertial and gravitational phenomena are always identical, it follows that light rays must be curved in passing through gravitational fields. Detailed analysis shows that this curvature should be exceedingly minute for the gravitational fields which we can study most easily. For light rays passing very close to the sun it should amount to about 1.7 seconds of arc. Before devising his General Theory, Einstein calculated that on the Newtonian theory of gravitation, assuming that light has mass, the corresponding displacement should be one-half this value.

The phenomenon has been tested at eclipses of the sun. As

seen from the earth, certain stars will appear to be in the same region of the sky as the sun at a particular epoch. When apparently very close to the sun they will be observed only during a total solar eclipse. On various expeditions, notably those in connection with the total eclipse of 29th May, 1919, measurements were made of the positions of such stars. Although it was widely believed that the results then obtained indicated that there was a displacement which was closer to that derived from General Relativity than to that derived from the Newtonian theory, the measurements were not much larger than the probable experimental errors. Consequently, this test is not quite so conclusive as the others.

Einstein has drawn attention to an important theoretical consequence of the bending of light rays in a gravitational field. According to the General Theory of Relativity, it appears that the law of the constancy of the velocity of light *in vacuo*, which constitutes one of the two fundamental assumptions of the Special Theory, does not apply universally. We again see that the Special Theory of Relativity is valid only as long as we can neglect the influence of gravitational fields on the phenomena considered. Nevertheless, in practice, owing to the weakness of most gravitational fields and the minute influence they exert on the transmission of light, it follows that the Special Theory is appropriate to a wide class of physical phenomena.

The Special Theory of Relativity differs in several important respects from the General Theory. It has received a considerable degree of experimental support, whereas the General Theory has a far less impressive list of crucial experimental tests to its credit. Moreover, the Special Theory has a clearer experimental foundation. We begin with measuring rods, clocks and light signals and deduce the geometry of the space-time of Minkowski involving numbers whose physical significance is clear. On the other hand, in developing the General Theory of Relativity, we begin with an abstract geometrical assumption. The outstanding achievement of the General Theory is the reduction of gravitation to geometry, but there is an ambiguity latent in this method. Space-time is curved in the neighbourhood of material masses, but it is not clear whether the presence of matter causes the curvature of

space-time or whether this curvature is itself responsible for the existence of matter. Indeed, in developing the theory this ambiguity continually arises. The expressions for the energy and momentum of a given material system depend on certain numbers characterizing the structure of space-time, but these numbers in turn depend on the distribution of matter contemplated.

Furthermore, although the Special Theory of Relativity does not account for electromagnetic phenomena, it explains many of their properties. General Relativity, however, tells us nothing about electromagnetism. In Einstein's space-time continuum gravitational forces are absorbed in the geometry, but electromagnetic forces are quite unaffected. Various attempts have been made to generalize the geometry of space-time so as to produce a unified field theory incorporating both gravitational and electromagnetic forces. However, none of these theories have produced any crucial experimental tests, and as they depend on certain arbitrary assumptions there is no conclusive evidence in their favour.

The philosophical consequences of the General Theory of Relativity are perhaps more striking than the experiential tests. As Bishop Barnes has reminded us, "The astonishing thing about Einstein's equations is that they appear to have come out of nothing." We have assumed that the laws of nature must be capable of expression in a form which is invariant for all possible transformations of the space-time co-ordinates and also that the geometry of space-time is Riemannian. From this exiguous basis, formulae of gravitation more accurate than those of Newton have been derived. As Barnes points out: "The conclusion seems to be irresistible that such laws of nature as the principle of conservation of energy, the principle of conservation of momentum and the law of gravitation are *necessary consequences of our modes of measurement*. They are, in fact, elaborately disguised identities which could have been predicted *a priori* by a being of sufficiently powerful analytical insight who fully understood all that is implied in the way in which we measure space-time intervals."

CHAPTER V

WORLD-MODELS (I)

OUR direct knowledge of the universe is confined to a limited region of space and time. In order to obtain some picture of the universe as a whole, we must construct a world-model which will reproduce satisfactorily the properties of this observable region. In the second chapter we discovered that our interpretation of the observations of the extra-galactic nebulae depended on our principles of measurement. Consequently, our model should provide an adequate background for the general laws of physics, including the theory of physical measurement, in terms of which our observations are interpreted. As the theory of relativity is essentially a theory of physical measurement, it is not surprising that it has played a prominent part in cosmology. Indeed, modern theoretical cosmology originated in a famous pioneer memoir, published by Einstein in 1917, depending on certain special assumptions about the universe as a whole.

In the Middle Ages the universe was regarded as finite, with the earth at its centre. This idea was abandoned during the Scientific Renaissance, and the universe came to be pictured as consisting of an indefinitely large number of stars scattered throughout infinite Euclidean space. This conception appeared to be a necessary consequence of the theory of gravitation; for, as Newton pointed out, a finite material universe in infinite space would tend to concentrate into one massive lump. In the first of his *Four Letters to Bentley*, written in 1692, he said:

“It seems to me, that if the matter of our sun and planets and all the matter of the universe, were evenly scattered throughout all the heavens, and every particle had an innate gravity towards all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of this space would, by its gravity, tend towards all the matter on the inside, and, by consequence, fall down into

the middle of the whole space, and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space."

During the nineteenth century certain difficulties in this conception were revealed. In 1826, Olbers suggested that an infinity of stars of the same average brightness of the sun would cause the sky to be infinitely bright, or at least to appear as bright as the sun in each direction. This difficulty can be overcome if we make special assumptions, e.g. that the stars in the more remote regions of space are much fainter than those in our neighbourhood, or that light is gradually absorbed during its passage through space.

A further objection was pointed out by Seeliger in 1895. He maintained that there could not be a uniform distribution of matter in the universe obeying the Newtonian inverse square law of gravitation. For on this theory the number of lines of force which come from infinity and end in a mass m is proportional to m . Now suppose the universe contains a uniform distribution of matter of density ρ . Then a sphere of volume V will enclose a mass ρV , and so the number of lines of force passing through the surface of the sphere will be proportional to ρV . But the area of the surface of the sphere is proportional to R^2 and its volume to R^3 . Therefore the number of lines of force which pass through a unit area of the sphere will be proportional to ρR , and so the intensity of the gravitational field at the surface of the sphere will increase indefinitely as the sphere expands. This is impossible, since the centre of the sphere can be anywhere we choose. Seeliger suggested that this difficulty could be overcome by an *ad hoc* amendment to the Newtonian law, which is negligible except when immense distances are involved.

An interesting attempt to reconcile the idea of an infinite distribution of stars with Newtonian gravitation was made some years ago by the Swedish astronomer Charlier in terms of a model originally suggested by the eighteenth-century

mathematician Lambert. In this model a certain number of stars form a nebula, or system of the first order, so many nebulae form a super-nebula, or system of the second order, and so on indefinitely. Charlier showed how such a model could be constructed, within the framework of classical theory, without encountering the difficulties indicated by Olbers and Seeliger. Hubble, however, has shown that despite local clustering effects there is no observational evidence so far for the existence of more than one system of the second order.

In considering the structure of the universe as a whole, Einstein assumed that to a first approximation the irregularities in the distribution of matter can be neglected. Moreover, he noted that the largest velocities then assigned to stars and nebulae were very small compared with the velocity of light. In 1917, of course, the immense red-shifts of the farther nebulae were unknown. Einstein, therefore, considered a model of the universe in which matter was distributed in a uniform and continuous manner, the relative motion of the various parts being negligible. Following Seeliger, he found it impossible to regard the system as filling the whole of Euclidean space. Also he could not regard the universe as an island in infinite space. For, by applying a well-known theorem of Boltzmann relating the densities at various points of space in which a distribution of particles is moving at random, he showed that zero density at the boundary would necessitate zero density at all points inside.

Thus it appeared to Einstein that the universe as a whole could be neither infinite nor have a finite boundary. Hence space as a whole could not be Euclidean. We have already considered a space-time in which the curvature varies from one region to another, but in constructing a model of the universe as a whole local irregularities are neglected. Thus the space-time of the model universe must have the same properties at all points. In the classical picture time and space are distinct, time being infinite in duration and space Euclidean. In devising an alternative model Einstein retained this world-wide separation of time and space, despite the fusion of the two concepts in General Relativity, but he assumed that space as a whole was of the type known as spherical.

There are three types of space which have the same properties at all points: Euclidean or 'flat' space, Lobatchewskian or hyperbolic space, and spherical space, but of these only spherical space is finite. This space is the three-dimensional equivalent of the surface of a sphere. This surface is unbounded, in the sense that we can move over it without coming to any edge, as we should do in the case of a disc. On the other hand, it has a finite area. In a similar fashion, spherical space is unbounded but has a finite volume. Moreover, we can take any point on the surface of the earth as a centre on the surface. Similarly, spherical space is symmetrically disposed with respect to every constituent point. Consequently, a uniform distribution of matter in spherical space would not automatically tend to cluster around any particular point as centre because every point would be central.

In order to obtain a model of this type in accordance with General Relativity, Einstein found it convenient to modify his law of gravitation slightly by introducing an additional term involving an entirely new constant, λ , known as the cosmical, or cosmological, constant. In considering local gravitational problems, for example the motion of the planets around the sun, this term can be neglected. Einstein assumed that it was significant only when considering the universe as a whole. In constructing his world-model, Einstein found that λ was equal to the reciprocal of the square of the radius. He also discovered a simple relation between the total mass M of his model and its radius R , *viz.*

$$GM = \frac{1}{2}\pi c^2 R,$$

where G is the constant of gravitation and c the velocity of light. Hence the total extent of space would be determined by the quantity of matter in the world. For example, by doubling the amount of matter we would automatically double the radius. Incidentally, since the volume of spherical space is $2\pi^2 R^3$, if the average density of matter in space is ρ , then

$$R^2 = c^2/4\pi\rho G,$$

indicating that the greater the density the smaller the radius. Thus in Einstein's model the curvature of the universe as a

whole depends on the amount of matter present. It follows that on this theory there would be a definite upper limit to the size of a sphere of any particular substance. For example, a sphere of water of radius about 200 million miles would fill the world. Moreover, since an increase in the average density of matter in the model would reduce the radius, we find that Einstein's universe contains as much matter as it can possibly hold for its spatial size.

Another interesting feature of the Einstein universe is that in principle it could be circumnavigated by a ray of light, although strictly speaking it contains no radiative phenomena. Various crude estimates of the time this would take have been made, but it appears unlikely that it would be less than 1,000 million years and probably several times as long. After such a time lapse the rays of light from any star would converge again at or near the starting point. Corresponding to the original star we should then see a ghost-star occupying the position where that star had been so many thousands of millions of years ago when these light rays were first emitted. In the actual universe it is unlikely that the rays would converge with sufficient accuracy. Nevertheless it is interesting to consider the possibility that some of the stars and nebulae which we see may after all be only optical ghosts.

Various objections have been brought forward against the Einstein model. First, absolute space and time are restored for the universe as a whole. There is a given radius of the world and a cosmic time, so that relativity is reduced to a local phenomenon. To this objection Eddington has replied that "the relativity theory is not concerned to deny the possibility of an absolute time, but to deny that it is concerned in any experimental knowledge yet found; and it need not perturb us if the conception of absolute time turns up in a new form in a theory of phenomena on a cosmical scale, as to which no experimental knowledge is yet available. Just as each limited observer has his own particular separation of space and time, so a being co-extensive with the world might well have a special separation of space and time natural to him. It is the time for this being that is here dignified by the title of 'absolute'."

Second, in Einstein's revised theory of world-gravitation the

total amount of matter in a world of given radius is determined by the law of gravitation. Eddington, writing on this in 1920, said, "Some mechanism seems to be needed, whereby either gravitation creates matter, or all the matter in the world conspires to define a law of gravitation." He then pointed out that, although he found this "rather bewildering", it was welcomed by those who follow Mach. "In the philosophy of Mach a world without *matter* is unthinkable. Matter in Mach's philosophy is not merely required as a test body to display properties of something already there . . . it is an essential factor in causing those properties which it is able to display. Inertia, for example, would not appear by the insertion of one test body in the world; in some way the presence of other matter is a necessary condition. It will be seen how welcome to such a philosophy is the theory that space and the inertial frame come into being with matter, and grow as it grows."

On the other hand, with regard to the Newtonian concept of absolute rotation, Eddington admitted that Einstein's plenum does in fact provide a world-wide inertial frame, with respect to which it can be measured. Nevertheless, Eddington believed that Einstein attributed too important a role to matter, for in his universe it appears that not only the metrical properties, as in General Relativity, but the very existence of space depends upon the existence of matter. Eddington preferred to regard matter as a manifestation of the 'structure' of space-time.

A further theoretical objection, thought to be decisive, was discovered by Eddington in 1930. Einstein's world is unstable, and if it experiences a minute disturbance it will either tend to expand or contract indefinitely. Eddington therefore suggested that Einstein's model could no longer be regarded as the approximate form of the physical world when 'smoothed-out' by the statistical averaging of observations on the distribution of matter. Instead, he suggested that it represented the initial state of the universe in the remote past. Of course, if the physical world actually began in this form, there could have been no *external* physical means of disturbing it. Eddington concluded that the initial disturbance must have been due to an inherent tendency in matter to change in some way. It is not necessary to adopt this line of argument, however, in

order to reject the Einstein model. Owing to local departures from strict homogeneity, the actual world cannot be *exactly* of Einstein's form. The instability of Einstein's universe indicates that, in general, a system which is nearly, but not exactly, of this form will tend to depart further and further from it with lapse of time. Consequently, Einstein's model cannot permanently represent the smoothed-out universe of nebulae. It is possible, of course, that the actual universe was much closer to this model in the remote past than it is now.¹

Shortly after Einstein published his original memoir on cosmology in 1917, de Sitter constructed an alternative static world-model, which satisfied the same laws of world-gravitation. In this model, unlike Einstein's, space-time has an intrinsic structure of its own, independent of the presence of matter. Indeed there is, strictly speaking, no matter nor radiation. Nebulae, if introduced into such a model, must therefore be considered as 'test particles', having no influence on the model as a whole. However, whereas a test particle in Einstein's universe will remain at rest if it has no initial motion, a similar particle introduced into de Sitter's world will immediately acquire an ever-increasing velocity of recession from the observer. Moreover, in de Sitter's model, space-time is 'hyperbolic'. There is no absolute time, and each observer will perceive an horizon at which time will appear to him to stand still, as at the Mad Hatter's tea-party where it was always six o'clock. This phenomenon, of course, is only apparent, like the rainbow. At any point on the (relative) horizon the time-flux experienced by an observer there will be the same as at the original observer. Thus in de Sitter's world there will be an *apparent* slowing-down of distant atomic vibrations, if these keep standard time. Consequently the radiation from a distant nebula will appear to be shifted towards the red, due to the increase in wavelength corresponding to the decrease in vibrational frequency. This effect, of course, will be supplemented by the Doppler effect, due to the relative recession of the nebula regarded as a test particle.

¹ It has even been suggested that the Einstein configuration was an unstable equilibrium state through which the universe slowly passed in expanding from an initial 'explosive' phase of small volume and high density.

It is clear that at best de Sitter's world, like Einstein's, can be regarded only as a limiting form of the real world. In the Einstein world there is the greatest possible concentration of matter without motion. In the de Sitter world there is motion but no matter. However, with the aid of a special hypothesis due to Weyl, it can be shown that a nebula introduced into the de Sitter world will exhibit a red-shift proportional to its distance, similar to Hubble's observational law. Robertson has remarked: "We should, of course, expect that any universe which expands without limit¹ will approach the empty de Sitter case, and that its ultimate fate is a state in which each physical unit—perhaps each nebula or intimate group of nebulae—is the only thing which exists within its own observable universe."

The models of Einstein and de Sitter are static solutions of Einstein's modified gravitational equations for a world-wide homogeneous system. They both involve a positive cosmological constant λ , determining the curvature of space. If this constant is zero, we obtain a third model in classical infinite Euclidean space. This model is empty, the space-time being that of Special Relativity.

It has been shown that these are the only possible *static* world-models based on Einstein's theory. In 1922, Friedmann, a Russian meteorologist, broke new ground by investigating non-static solutions of Einstein's field equations, in which the radius of curvature of space varies with time. This possibility had already been envisaged, in a general sense, by Clifford in the eighties. He suggested that universal space might have a constant curvature which "may change as a whole with the time. In this way our geometry based on the sameness of space would still hold good for all parts of space, but the change of curvature might produce in space a succession of apparent physical changes." Friedmann restricted his investigations to spaces of positive curvature, i.e. closed or finite spaces, but allowed the cosmological constant to be positive, zero or negative.

Five years later, in 1927, a more vivid account of the sub-

¹ That is to say, with continual acceleration. In de Sitter's model any test-particle will ultimately exceed the velocity of light, and so will escape from the observer's realm of possible observation.

ject was given independently by the Belgian mathematician Abbé Lemaître, who worked out the astronomical consequences in considerable detail. He, too, considered only spaces of positive curvature, and it was not until 1931 that the German astronomer Heckmann pointed out that both the curvature and the cosmical constant could be positive, zero or negative.

The investigations of Friedmann and Lemaître, although studied and extended by Robertson and other mathematicians, were neglected by most astronomers until the spring of 1930 when, following Hubble's publication of the law correlating the distances and red-shifts of the extra-galactic nebulae, attention was directed to the construction of expanding world-models. Eddington then discovered the 'instability' of the Einstein universe, and was led to regard the real universe as a system expanding from the Einstein model as an initial state of unstable world-equilibrium between two opposing forces, namely world-gravitation and 'cosmical repulsion' depending on positive λ .¹

The necessity for introducing the cosmical constant λ into the gravitational equations of General Relativity was originally due, as we have seen, to Einstein's assumption that *all* stellar velocities are very small compared with that of light. Once this assumption was no longer required, the theoretical necessity for introducing λ became much less obvious, and in 1932 Einstein and de Sitter, in a joint memoir, argued that the original objections of 1917 to a world-model of finite density in Euclidean space no longer applied if space could be regarded as expanding. As they pointed out: "There is no direct observational evidence for the curvature, the only directly observed data being the mean density and the expansion, which latter proves that the actual universe corresponds to the non-statical case. It is therefore clear that from the direct data of observation we can derive neither the sign nor the value of the curvature, and the question arises whether it is possible to represent the observed facts without introducing the curvature at all.

¹ According to this conception, the universe is now expanding more rapidly than in the past, and its age is consequently greater than Hubble's 1,800 million years. Eddington estimated an age of 90,000 million years, in a memoir published in 1944.

Historically the term containing the 'cosmological constant λ ' was introduced into the field equations in order to enable us to account theoretically for the existence of a finite mean density in a static universe. It now appears that in the dynamical case this end can be reached without the introduction of λ ." Einstein and de Sitter therefore constructed a homogeneous world-model of finite density, subject to the field equations of General Relativity in expanding Euclidean space. They found that, with the observed rate of expansion of approximately 500 kilometres per second per megaparsec (a million parsecs), their model predicted a smoothed-out density of approximately 4×10^{-28} grammes per cubic centimetre. Although this density "may perhaps be on the high side, it is certainly of the correct order of magnitude and we must conclude that at the present time it is possible to represent the facts without assuming a curvature of three-dimensional space".

However, in 1935, Milne pointed out that this apparently simple world-model has a most curious property. 'To any observer in it the total number of nebulae which he could observe with a telescope of unlimited power would be directly proportional to the epoch of observation, so that with the march of time new nebulae would continually 'swim into his ken' and he could thus actually witness the creative process of the world out of nothing. As Milne explained, "The mathematics is responsible for this creation of matter." It is asked to produce a finite number of observable objects in a Euclidean space, such that each is central in the field of the remainder. "It only achieves this object by *creating* fresh particles beyond each given particle as fast as they are required; and it brings them to birth with the velocity of light. The frontier of limiting range of observability moves onward with the speed of light; it contains always the particles just being created, and it leaves in its wake a spray of decelerating newly created particles." Milne also stressed the corresponding results for the Friedmann-Lemaître universes. In particular, if the cosmical constant is positive, we find a corresponding observable annihilation of matter, due to its disappearance from the field of possible observation, as it is accelerated beyond the limiting velocity of light *in vacuo*.

In the three static universes of General Relativity the cosmical constant is related to the radius of curvature, R ; in the Einstein model, λ is equal to the reciprocal of R^2 , and in the de Sitter model, λ is equal to $3/R^2$. In the only other static model, that with the space-time of Special Relativity, R is infinite and λ is zero. Hence in each of these models λ has a definite physical interpretation. It is a quantity intimately related to the radius of curvature of the world, and so provides a natural unit of *length*, except in the Special Relativity case where it does not appear.

On the other hand, in those models in which the radius varies with the epoch, there is no fixed relation between the constant λ and the variable radius. The different possibilities were conveniently tabulated in 1932 by de Sitter as follows:

λ	CURVATURE		
	<i>Negative</i>	<i>Zero</i>	<i>Positive</i>
<i>Negative</i> ..	Oscillating	Oscillating	Oscillating
<i>Zero</i> ..	Expanding I	Expanding I	Oscillating
<i>Positive</i> ..	Expanding I	Expanding I	{ Oscillating Expanding I Expanding II }

In the oscillating models the 'radius' increases from zero to a certain maximum value, which varies according to the model, and then decreases again to zero, the period of oscillation varying from one to another. In the expanding models of the first type (I), the radius continually increases from zero at some definite epoch and tends to infinity after an infinite lapse of time. The expanding models of the second type (II) behave in the same way, except that at the initial epoch when expansion begins the radius has a definite non-zero value, which is

different for each model. Incidentally, when the curvature of space is zero, so that the 'radius' is infinite, expansion simply means that the distances between all 'fixed' points embedded in the spatial plenum increase with change of epoch. Thus the plenum is like a spotted block of rubber which is being continually distended in all directions.

We see that General Relativity presents us with an almost embarrassing plethora of 'homogeneous' expanding systems. Moreover, it provides no *conclusive* theoretical explanation of why the universe is expanding and not contracting.¹ Even if the *available* observational data were far more precise, we should still be unable to pick out a unique relativistic model. For our data, in essence, isolate only *two* unknowns, viz. the mean density and the rate of expansion (if we interpret the red-shifts as due to the Doppler effect). But in our relativistic models we find *three* unknowns, viz. λ , the sign of the curvature, and its magnitude at the present epoch, i.e. the scale of the model. Hence it would appear that, in principle, any one of these three could be chosen arbitrarily. Consequently, as was indicated at the end of the second chapter, it is not surprising that controversy has raged for the last two decades between the respective protagonists of the various world-models.

Here we can only barely indicate the rival interpretations which have been attributed to the available data. For example, Hubble, who is the leading observational authority, and has the facilities of the Mount Wilson observatory at his disposal, inclines to the view that the data fit most convincingly into a *static* model in classical Euclidean space. However, as usual, there is a price to pay for this apparent simplicity: Hubble is completely baffled by the notorious red-shifts, which he can only attribute to a mysterious new principle of nature. To quote his own words:

"If nebulae are uniformly distributed through a non-expanding universe in which red-shifts are not primarily velocity-shifts, then the numbers should be proportional to the volumes, and the surveys should conform (and actually do conform) with the relation

¹ It has been shown that expansion *may* result if condensations begin to form in the Einstein universe.

$$\log_{10} N = 0.6 m_c + \text{constant},$$

where m_c is the limiting faintness expressed as a magnitude, corrected for local obscuration and for the energy-effects required by the mere presence of red-shifts. (The derivation is as follows. For uniform distribution, numbers of nebulae N are proportional to volumes of space, and consequently to the cubes of the limiting distances r to which counts are carried. Hence $\log_{10} N = 3 \log_{10} r + \text{constant}$. Among the objects of the same candle-power the distances are proportional to the inverse square of the apparent luminosity L . Hence $\log_{10} r = \text{constant} - 0.5 \log_{10} L$. Apparent magnitudes m measure apparent luminosities on a logarithmic scale. By definition, $m = \text{constant} - 2.5 \log_{10} L$. Hence $\log_{10} r = 0.2 m + \text{constant}$, and $\log_{10} N = 0.6 m_c + \text{constant}$.) The corresponding relation for a homogeneous expanding universe, obeying the relativistic laws of gravitation, is

$$\log_{10} N = 0.6 (m_c - d\lambda/\lambda + C_v) + \text{constant},$$

where $d\lambda/\lambda$ is the recession factor and C_v is the effect of spatial curvature."

In order to get agreement with observations, Hubble claims that we require that

$$C_v = d\lambda/\lambda,$$

and remarks: "To the observer the procedure seems artificial. He has counted the nebulae to various limits, applied only the corrections that are necessarily required (energy-corrections), and derived the quite plausible result of uniform distribution. Now, in testing the relativistic theory, he introduces a new postulate, namely recession of the nebulae, and it leads to discrepancies. Therefore he adds still another postulate, namely spatial curvature, in order to compensate the discrepancies introduced by the first. The accumulation of assumptions is uneconomical, and the justification must be sought in the general background of knowledge."

"Therefore," continues Hubble, "the expanding universe can be saved by introducing a sufficient amount of spatial

curvature. The plausible values are narrowly limited, and they indicate a radius of curvature that is positive and comparatively small. In fact the radius, about 470 million light-years, is a trifle less than the range of the 100-inch reflector for normal nebulae." It follows that, *if* the universe is expanding and *if* the calculations made by Hubble (with the assistance of the relativity expert Tolman) are justified, space is closed and has a 'remarkably small' radius of curvature, so that "a large fraction of the universe, perhaps a quarter, can be explored with existing telescopes". The volume of this universe would be $2\pi^2 R^3$, where R is the radius of curvature, or about 2×10^{27} cubic light-years, and it would contain only about 400 million nebulae.

Having determined the nature of the curvature and its magnitude for this model, Hubble then determines the corresponding range of values of the cosmological constant λ , assuming that neither the mean density in the universe nor the mean pressure are less than zero. The range is found to be small and the approximate value of the constant is 4.5×10^{-18} (years) $^{-2}$, i.e. positive and slightly greater than its critical value in the unstable Einstein universe. Thus, according to de Sitter's classification, Hubble's expanding model is of the first kind (I); expansion began from an arbitrarily small volume about 1,000 million years ago and will continue for an infinite future time. At first expansion was rapid, but it is continually slowing down. "At each moment the model is homogeneous, the contents are uniformly distributed, but as time goes on the mean density diminishes, the average distance between neighbouring nebulae increases. Eventually a state of complete isolation will be reached."

Hubble points out some disturbing features in this model: the very small scale both in space and time, so that with existing telescopes we explore at least a third of the 'age' and about a quarter of the volume; also, since the formulae yield a relation between λ , the curvature and the contents of space, the mean density is of the order of a thousand times that which can be accounted for by the nebulae observed. If such excess matter exists, it must be scattered between the nebulae, and would probably be in the form of dust or ionized gas. However, such a

medium would absorb light to a far greater degree than Hubble believes possible.¹ On the other hand, none of these disturbing features characterize Hubble's alternative non-expanding model. Here the difficulty, as we have already remarked, is to account for the characteristic spectral shifts.

Hubble's detailed procedure has been criticized by several authorities. About twelve years ago Tolman pointed out that Hubble and Humason had based their calculations of nebular distances on two questionable hypotheses, namely:

(1) That the average *absolute* luminosity (or magnitude) of the more distant nebulae so far observed is equal to that of the nearer nebulae;

(2) That the *apparent* luminosity of the nebulae, making due allowance for the effect of the red-shift, is proportional to the inverse square of their distances, as for stationary objects in Euclidean space.

Until recently it appeared that the first assumption might not be very serious. For, as the distance of the farthest nebulae so far observed is of the order of 5×10^8 light-years, this assumption is equivalent to postulating that the intrinsic brightness of the average nebula has not changed appreciably in the past 500 million years. For this we have some independent evidence, although it is by no means conclusive. From fossil records it is clear that life has existed on the earth's surface for at least an equivalent period. Consequently, the continual outpouring of radiation from the sun during this time cannot have varied greatly. If the brightness of the average nebula is due to an appreciable extent to stars similar to our sun, it is possible that its intrinsic luminosity has also been sensibly constant over the past 500 million years.

Recently, however, Stebbins and Whitford have found that the 'colour-indices' of certain nebulae appear to increase with red-shift, so that the more distant nebulae appear redder than is indicated merely by their red-shifts on the assumption of standard absolute luminosity. This phenomenon might be accounted for if in the past these nebulae contained many red

¹ It is not unlikely that internebular space contains diffuse matter, but its average density is probably much less (e.g. a thousand times) than that of interstellar matter in the Galaxy.

super-giant stars which have since faded. Whatever the explanation, the effect on the determination of distance would appear to be serious.

The second assumption of Hubble and Humason has long appeared to be troublesome, for the correction that would be required if space were curved, either spherically or hyperbolically, would depend on the square of the ratio of the distance of the nebulae concerned to the radius of curvature of the space. That this correction is not large over the present range of observations is by no means certain. Indeed, the whole problem of distance bristles with complications, both practical and theoretical, as was shown in a series of important researches about fifteen years ago by the Edinburgh School (Whittaker, Walker, Ruse, McCrea and McVittie), who considered in great detail the problem of interpreting spatial distance in General Relativity and relating it to empirical determinations based on the apparent luminosities, etc., of the nebulae.

In an attempt to differentiate as far as possible between the 'empirical' and the 'theoretical', McVittie showed in 1939 that "the information we can hope to secure from observation is not as extensive as recent researches on the structure of the universe might lead us to suppose". He claimed that the sign of the curvature and its magnitude cannot be determined independently. "To advance further we must again appeal to theory." Instead of employing Hubble's assumption that nebular spectra, on the average, are similar to the spectrum of the sun, McVittie took as his 'arbitrary datum' the average mass of a nebula and invoked Einstein's field equations relating the pressure and density of matter in the universe to its geometry. Assuming that the average nebula has a mass of $4 \times 10^{43} \pm 1$ grms. and that there are no invisible nebulae, he obtained a hyperbolic space of radius 1.5×10^9 parsecs and a mean density of from 10^{-28} to 10^{-30} grms. per cubic centimetre. More generally McVittie concluded: "That space is hyperbolic seems nearly certain. . . . The average density is certainly not greater than 10^{-27} grms. per cubic centimetre and is more probably of the order of 10^{-29} grms. per cubic centimetre. The effective temperature of the nebular radiation consonant with these results lies between 5000° and 7500° ."

In 1948, however, Bondi pointed out that McVittie's results depended on the particular formula, due originally to Hubble, which he had employed for calculating the total amount of radiation received from a nebula from the amount which directly affects the photographic plate. The relation between these two quantities varies with spectral type and is affected by the red-shift. Unfortunately, it has not been possible so far to obtain spectra of the more distant nebulae. Bondi has shown that McVittie's results are very sensitive to the mathematical form of the relation and claims that a more general formula than Hubble's, when applied to McVittie's analysis, yields a *closed* universe, although of considerably greater radius than that of Hubble's closed model.

We can now appreciate more fully the point made at the end of the second chapter: that, in interpreting the raw data of observation, theory and observation are inextricably interwoven, and until a comprehensive observational programme has been undertaken with the new 200-inch telescope at Mount Palomar, due to be in operational use in 1949, little further progress is likely to be made along the lines discussed in this chapter. But before passing on to consider world-models of a somewhat different character, tribute must be paid to the masterly observational technique of the American astronomers, particularly Slipher, Hubble and Humason. Our present knowledge is based on their remarkable photography and delicate measurement of almost imperceptible data.

CHAPTER VI

WORLD-MODELS (II)

IN constructing the world-models so far described, the primary object has been to reproduce the general features of the observable region. We have seen that this is not possible without some theoretical presuppositions. The present confusing situation, in which observations appear to be insufficient to isolate a unique model, is due in great part to disagreement among the experts concerning the best choice of these presuppositions. We have, however, already emphasized the peculiar dual object of a world-model. Unlike all other types of physical model which are concerned with only a part of the physical universe, a *world*-model must not only reproduce satisfactorily the relevant general features of observation, but must also provide its own laws of nature, by which this agreement with observation can be judged. In other words, a truly satisfactory world-model must provide an adequate background for the general laws of physics, including the theory of physical measurement, in terms of which our observations are interpreted.

Eddington and Milne have constructed world-models with this more fundamental object in view. In their models detailed comparison with the observed properties of stars and nebulae is secondary; the first object is to erect a steel framework of physical measurement, which can support the stone facing of observation and experiment. To change the metaphor, Eddington and Milne have each tried to put the horse before the cart. Thus, whereas in the pioneer investigations of Einstein and de Sitter, Friedmann and Lemaître, cosmology is derived from General Relativity as an extrapolation, in the more recent work of the two great British theoreticians, world-principles are axiomatic, and principles appropriate to local phenomena, e.g. General Relativity and other methods of analysing gravitating systems, are, or should be, derived therefrom.

The origin of Eddington's formal approach to cosmology

was Dirac's momentous discovery in 1928 that in the theory of the electron there arise mathematical expressions invariant under certain types of transformation, which could *not* be obtained by the mathematical method (tensor calculus) employed in General Relativity. As Eddington has since remarked, "We had claimed to have in the tensor calculus an ideal tool for dealing with all forms of invariance. . . . Why had this type of invariance eluded the ordinary tensor calculus? As C. G. Darwin put it, 'it is rather disconcerting to find that apparently something has slipped through the net'. . . . The failure of ordinary tensor calculus to include Dirac's type of invariance is due to the introduction, at an early stage, of a convention whose arbitrariness had already been noticed. The analytical theory of tensors had been applied to physics by identifying its basic vector¹ with a geometrical displacement. . . ."

It will be remembered that in the fourth chapter we stressed the characteristic difference in method between Einstein's Special and General Theories of Relativity. In the former he began with (ideal) physical processes, measurements of distance and epoch by means of light signals, rigid rods and clocks. It was found that, although different observers in uniform relative motion employed 'similar' apparatus, in general they assigned different lengths to the same object and different epochs to the same event. However, by combining spatial and temporal measures in a particular way, an *abstract* quantity was constructed which had the important property of invariance, i.e. the property of having the same value according to each observer moving uniformly with respect to a given initial observer.

In developing the General Theory of Relativity Einstein began with an algebraic device. He postulated the invariance with respect to all frames of reference (i.e. for all possible transformations) of a generalization of Minkowski's space-time. Although, following Minkowski's lead, he regarded space-time as a peculiar combination of space and time, "Henceforth, space by itself and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality"; nevertheless the space-

¹ Interval between two events at different places.

time of General Relativity is a kind of four-dimensional space and the basic vector, as Eddington remarks, is, in essence, a 'geometrical displacement'. In a letter to J. W. Dunne in 1928 Eddington wrote: "the 'going on of time' is not in Minkowski's world as it stands. My own feeling is that the 'becoming' is really there in the physical world, but is not formulated in the description of it in classical physics." The same criticism applies to the space-time of General Relativity.

In General Relativity, as in classical dynamics, time can usually be 'reversed' without affecting the fundamental equations. Now, in considering problems of 'local' gravitation, e.g. the motion of the planets round the sun, this does not matter, except possibly over exceedingly long periods of time. For example, if we could take a film of the motion of Mercury and the sun over hundreds, or even thousands, of years, and then were to run it backwards through a cinema projector, *the general effect* would be the same. When, however, we come to consider world-phenomena and contemplate processes over hundreds and thousands of millions of years, we must take into account the essential irreversibility of time. To use two excellent terms recently employed by Ubbelohde in an essay on *Time and Thermodynamics*, we must then distinguish between time as 'trend' and time as mere 'duration'.

The limitations of General Relativity are mainly due to the arbitrary features associated with the basic vector. The recent investigations of Eddington and Milne had their common origin here, although they diverged widely from the start. Whereas Milne has concentrated on the flux of time, Eddington developed new mathematical techniques for the study of world-invariants, in particular the so-called 'Constants of Nature'.

In 1922, following a pioneer suggestion made by Weyl in 1919, Eddington welcomed the possibility of a *finite* universe, as indicated by Einstein, because of the following physical argument. Among the constants of nature there is one, viz. the ratio of the electrical force to the gravitational force between two electrons, which is a very large pure number, of the order of 3×10^{42} . "It is difficult," Eddington wrote, "to account for the occurrence of a pure number (of order greatly different from

unity) in the scheme of things; but this difficulty would be removed if we could connect it with the number of particles in the world." Of course, this would be possible only if the number of particles were finite, as in the Einstein universe.

Two years previously, as already mentioned in the previous chapter, Eddington had found it "very hard to accept" Einstein's model, because the total amount of matter in it is determined by the law of gravitation. This result followed from Einstein's revised law of gravitation involving the cosmical constant λ . However, in 1932, Eddington pointed out that, although Einstein's original reason for introducing λ was not very convincing, and for some years "was looked on as a fancy addition rather than as an integral part of the theory", nevertheless as a result of "the great advance made by Prof. Weyl, in whose theory it plays an essential part . . . return to the earlier view is unthinkable. I would as soon think of returning to Newtonian theory as of dropping the cosmical constant."

In 1934 he wrote: "The cosmical constant expresses a relation of scale between two different types of phenomena.¹ So long as it is expressed by any number, however small, the relation remains recognized. But if it is expressed by zero the relation is broken. . . . It was a defect of Einstein's original theory, first remedied by H. Weyl, that it implied the existence of an absolute standard of length—a conception as foreign to the relativistic point of view as absolute motion, absolute simultaneity, absolute rotation, etc. To set $\lambda = 0$ implies a reversion to the imperfectly relativistic theory—a step which is no more to be thought of than a return to the Newtonian theory."

We thus see that, with the development of Eddington's ideas, the *raison d'être* of the cosmical constant was transformed so that, although the same symbol λ was retained, the concept denoted was quite different. Supporting his predisposition towards a world-model containing a finite number of particles in a finite spherical space, characterized by a cosmical constant, with a new and beautiful mathematical technique suggested by Dirac's discovery, Eddington embarked upon his boldest flight of thought. This was illustrated in two impressive books: *The Relativity Theory of Protons and Electrons*, published in

¹ For example, the radius of the electron and the radius of the universe.

1936, and *Fundamental Theory*, published posthumously in 1946.

It is impossible to convey within the scope of this essay more than a faint glimmer of this extraordinary mental achievement, and the time has not yet come to pass final judgment upon it. In a recent review of Eddington's last book, Chapman writes: "Whereas Eddington was regarded among astronomers throughout the world with immense admiration and respect, his work on 'fundamental theory' brought him obloquy, scoffing and suspicion from the theoretical physicists. At best he was pronounced incomprehensible, at worst he was accused of fudging his formulae (if not intentionally, yet nevertheless actually). The number of men with the great breadth of knowledge needed to assess his work is not great. . . . Personally, I cannot claim to have the knowledge needed to pronounce upon the matter; but my association with Eddington and acquaintance with his work in other fields leads me to expect that his work on 'fundamental theory' will prove well founded."

Although in its final form Eddington's theory is independent of General Relativity and is based on a new philosophy of physical measurement, the keystone of the whole structure is Einstein's idea that the world is finite. This principle guided Eddington in his initial choice of a problem which could be solved both by General Relativity and Quantum Mechanics, thus providing a bridge between these two great disciplines of modern theoretical physics, uniting the astronomical universe of nebulae with the atomic universe of protons and electrons. This problem concerned the equilibrium state of a radiationless self-contained system of a very large, but finite, number of particles. Regarding this problem as an exercise in cosmical physics, Eddington in his preliminary calculations adopted Einstein's pioneer solution of 1917, and took over the relation, quoted in the last chapter,

$$\frac{GM}{c^2} = \frac{1}{2} \pi R, \quad (1)$$

R being the radius and M the mass of the system, G the constant of gravitation and c the velocity of light. However, in

adopting this solution, Eddington interpreted M in a special way. He regarded it as being effectively equivalent to Nm_p , where m_p is the mass of the proton, or nucleus of the hydrogen atom, and N is the total number of protons in the system. For the whole system to be electrically neutral, he assumed that it also contained N electrons, but as the mass of a proton is about 1,840 times that of an electron, the contribution of the electrons to M can be neglected. An Einstein universe of this type Eddington eventually called a 'uranoid'.

In Quantum Mechanics, detailed discussion of which lies outside the scope of this essay, a radiationless steady system is said to be in its 'ground state', i.e. state of lowest possible energy. There is a fundamental 'exclusion principle' due to Pauli, which asserts that in any atomic system there is only one particle at any given 'energy level'. Therefore, in solving the problem of the uranus as an exercise in atomic physics, Eddington assumed that the N protons occupied the N states of 'lowest energy'. He then found the total energy of the uranus in terms of N and R and the constants of atomic physics. As we have already mentioned in Chapter IV, Einstein had shown that the mass M of a system whose total energy is E is given by the formula $E = Mc^2$. Consequently, Eddington's quantum calculation led to a second formula for the mass of the uranus. By combining it with the first relation, he obtained two equations for the mass and radius of the uranus in terms of the fundamental constants of nature.

This calculation was continually revised by Eddington and was eventually based on a new type of relativity theory, but in its crude pioneer form it was not too difficult for others to understand. The essential point in the original method, devised in 1931, was the idea that the mass of the electron was determined by its charge and by the existence of all the other particles in the uranus. Eddington derived the relation,¹

$$\frac{e^2}{m_e c^2} = \frac{R}{\sqrt{N}}, \quad (2)$$

¹ The expression on the left is approximately the classical 'radius of the electron'.

where m_e is the mass and e the charge on the electron. Combining this with Eddington's form of (1), viz.,

$$\frac{GNm_p}{c^3} = \frac{1}{2}\pi R, \quad (3)$$

we see that

$$\frac{GNm_p m_e}{c^3} = \frac{1}{2}\pi \sqrt{N},$$

whence

$$N = \frac{\pi^2 e^4}{4 G^2 m_p^2 m_e^2}. \quad (4)$$

Substituting for π the well-known value $3.14 \dots$, and for the other constants their centimetre-gramme-second values, viz. $G = 6.67 \times 10^{-8}$, $m_p = 1.67 \times 10^{-24}$, $m_e = 9.1 \times 10^{-28}$, $e = 4.77 \times 10^{-10}$, we find that the order of magnitude of N is 10^{79} . Consequently, the order of magnitude of M is 10^{55} grammes, or ten billion billion billion billion tons, and the corresponding value of R is immediately found to be of the order of a thousand million light-years.

This attempt to 'weigh' and measure the whole universe must be regarded as one of the most striking feats of modern relativistic cosmology. We can now appreciate perhaps more readily than our forefathers the genius of the Greek mathematician Archimedes who made a remarkable attempt to estimate the size of the universe more than two thousand years ago. His treatise *The Sand Reckoner* has been regarded until recently merely as an arithmetical *tour de force* in which, notwithstanding the inadequacy of the current methods of numerical notation, the great Syracusan succeeded in expressing the most enormous numbers ever likely to arise in a physical problem, in particular the number of grains of sand which would be required to fill a sphere equal in volume to the whole universe.

"There are some, King Gelon, who think that the number of the sand is infinite in multitude; and I mean by the sand not

only that which exists about Syracuse and the rest of Sicily but also that which is found in every region whether inhabited or uninhabited. Again there are some who, without regarding it as infinite, yet think that no number has been named which is great enough to exceed its multitude. . . . But I will try to show by means of geometrical proofs, which you will be able to follow, that of the numbers named by me and given in the work which I sent to Zeuxippus, some exceed . . . that of a mass equal in magnitude to the universe. Now you are aware that 'universe' is the name given by most astronomers to the sphere whose centre is the centre of the earth and whose radius is equal to the straight line between the centre of the sun and the centre of the earth. . . .

" . . . But Aristarchus of Samos brought out a book consisting of some hypotheses in which the premises lead to the result that the universe is many times greater than that now so called. His hypotheses are that the fixed stars and the sun remain unmoved, that the earth revolves about the sun in the circumference of a circle, the sun lying in the middle of the orbit, and that the sphere of the fixed stars, situated about the same centre as the sun, is so great that the circle in which he supposes the earth to revolve bears such a proportion to the distance of the fixed stars as the centre of the sphere bears to its surface. Now it is easy to see that this is impossible; for, since the centre of the sphere has no magnitude, we cannot conceive it to bear any ratio whatever to the surface of the sphere. We must however take Aristarchus to mean this: since we conceive the earth to be, as it were, the centre of the universe, the ratio which the earth bears to what we describe as the 'universe' is the same as the ratio which the sphere containing the circle in which he supposes the earth to revolve bears to the sphere of fixed stars. For he adapts the proofs of his results to a hypothesis of this kind, and in particular he appears to suppose the magnitude of the sphere in which he represents the earth as moving to be equal to what we call the 'universe'."

In this treatise Archimedes develops a powerful new technique for expressing the order of magnitude of large numbers and also performs some subtle calculations of various astronomical angles and distances, but in the light of the most

recent cosmological theories the outstanding feature of this great work is the apparent similarity in method between Archimedes' approach to the problem and our own. For, in assuming that "the ratio which the earth bears to what we describe as the 'universe' is the same as the ratio which the sphere containing the circle in which he supposes the earth to revolve bears to the sphere of fixed stars", Archimedes all but adopted the modern hypothesis of world-homogeneity. For him this would mean that the average density of matter in the 'universe' is equal to the average density obtained by 'smoothing out' the earth so as to fill the entire sphere on which its orbit lies. It is, therefore, of some interest to compare Archimedes' result with Eddington's, allowing for the tremendous superiority in our present data concerning astronomical distances and sizes compared with the vague and inadequate knowledge of Archimedes.

Archimedes assumed that a quantity of sand not greater than a poppy-seed contains not more than 10,000 grains; and that the diameter of a poppy-seed is not less than one-fortieth of a finger-breadth, i.e. about half a millimetre. He calculated that the number of grains of sand which could be contained in a sphere of the size of our 'universe' is not greater than 10^{51} . Hence, since the density of sand is about twice that of water, it is easy to calculate that the 'weight' of Archimedes' universe is of the order of 10^{44} grammes, compared with Eddington's 10^{55} . Indeed, Archimedes' universe is comparable in mass with a nebula, e.g. our own Galaxy. On the other hand, assuming that the diameter of the sphere of fixed stars is less than ten thousand times the diameter of the 'universe'—and with Archimedes' crude data the radius of this sphere thus corresponds to only about one light-year—he found that the total number of grains of sand which could be packed into this sphere is 10^{63} . Hence the corresponding mass is approximately of the same order as the mass of the Eddington universe.¹

Recently it was shown by the author that Eddington's order of magnitude for the mass of the universe can be obtained by a concise argument partly based on Newtonian mechanics.

¹ This result is, of course, purely an interesting and unexpected numerical coincidence.

In the latter, the gravitational energy of a homogeneous sphere of mass M and radius R is

$$V = \frac{3}{5} \frac{G M^2}{R},$$

where G is the constant of gravitation. In general, this is very much smaller than the inertial energy assigned by Einstein,

$$E = M c^2,$$

where c is the velocity of light. According to Mach's principle, the inertia of a body is due to the background influence, or 'pull', of the whole universe. According to Einstein's principle of equivalence, gravitation is akin to inertia. Thus the gravitational background influence of the whole universe should be equivalent to its inertial influence. Hence, despite the great numerical discrepancies between E and V for most physical bodies, it is suggested that, if it is legitimate to regard a certain homogeneous 'sphere' as a first approximation to the universe, then for such a sphere it is reasonable to assume that E and V are equal. It follows that the mass and radius must be related by the law

$$GM = k R,$$

where k is $5/3$, or approximately 1.67 . We immediately observe that this relation is identical with that characterizing the Einstein universe, except that in the latter k is $\pi/2$, or approximately 1.57 . If we assume that R is of the order of $1,000$ million light-years, as is indicated by all attempts to fit the observations into a spherical model, it follows that M is of the order of 10^{55} grammes, or 10^{79} protons or nucleons.

In Eddington's theory, the fundamental problem of whether the universe is finite or infinite is answered by exploring the consequences for physics of the assumption of finitude, rather than by an elaborate analysis of the present incomplete astronomical data. On this assumption, Eddington erected an imposing edifice of physical theory and claimed that a complete explana-

tion of the laws of physics could be obtained thereby. He believed that the properties of matter depended on the properties of space and time, and hence on our methods of measurement. In particular he regarded N , the number of particles in the universe, as a fundamental characteristic of physical measurement which could be determined exactly by a penetrating analysis of physical theory. Following the work of Dirac, Eddington developed a new type of tensor calculus, later supplemented by a new type of statistical relativity, in which the number N played a vital role. Instead of being merely an astronomical curiosity, its exact evaluation permitted the 'exact' evaluation of most of the fundamental constants of physics.

We have seen that the order of magnitude of N is 10^{79} . It is readily calculated that the ratio of the radius of the proton (or electron) to the radius of the uranoid is about one part in 10^{39} . Similarly, the ratio of the gravitational force to the electrical force between a proton and an electron is also about one part in 10^{39} . The number 10^{39} is approximately the square root of N , and this suggests that these ratios are determined by the number of elementary particles in the world.

Eddington discovered precise formulae correlating the number N with other fundamental constants. These formulae were deduced by means of an ingenious application of Heisenberg's principle concerning the impossibility of assigning exact values to both the position and momentum¹ of an elementary particle. According to this principle there is a range of uncertainty in each, the two ranges being related by means of Planck's quantum constant h . By the theory of errors we know that if we have N independent particles, each with an uncertainty of position R , the uncertainty of position of their mean centre is R/\sqrt{N} . If, instead of using one particle as an origin of reference, we use the mean of N independent particles, we obtain an origin whose uncertainty of position is R/\sqrt{N} . Hence, if we take the uranoid as a fundamental frame of reference, we obtain an origin whose uncertainty of position is expressible in terms of the radius of the uranoid and the number of elementary particles. This uncertainty is found to be of the same order of magnitude as the radius of an ele-

¹ Product of mass and velocity.

mentary particle.¹ According to Heisenberg's principle, the corresponding irreducible momentum can be calculated, and the corresponding mass is obtained by dividing this by the velocity of light c .

Elaborating these ideas with the aid of extremely intricate and detailed algebra, Eddington obtained a precise relation correlating the radius of, and number of particles in, the uranoid with Planck's constant and the mass of the proton. Also, he constructed an equation relating the masses of the proton and electron, and obtained a theoretical value of 1834.34 compared with an 'observed' value² of 1834.27.

These remarkable developments of Eddington's theory were made possible by his *exact* evaluation of the number N . He interpreted this number as the number in spherical space of possible types of waves which, in accordance with recent atomic theory, can be associated with elementary particles. This number he ultimately evaluated as $3/2 \times 136 \times 2^{256}$, which is of the order of 10^{79} . Moreover, by adopting three standard theoretical quantities, the velocity of light, the Faraday and Rydberg constants for hydrogen, as initial data, he claimed to have calculated correctly to one part in five thousand the values of an impressive number of fundamental constants, including the constant of gravitation, Planck's constant h , and the ratio of the masses of the proton and electron.

Three fundamental characteristics of Eddington's work must be emphasized. First, unlike the exponents of current orthodox theory, Eddington did not assume that atomic phenomena occur in a vacuum. Instead, he invoked the presence of the whole material universe as a necessary background. Secondly, Eddington claimed that many of the results of experiment and observation can be deduced theoretically from the way in which we measure phenomena, and are thus true *a priori*. This claim has been strongly contested, and although it is extremely difficult to follow Eddington's arguments, it is equally difficult to prove conclusively that he was wrong.

¹ See formula (2) above. We have now interpreted the right-hand side of this equation.

² This value is the standard observed value, 1847.6, 'corrected' by the factor 136/137.

Whittaker has suggested that Eddington's work is similar to that of Archimedes in evaluating the number π , denoting the ratio of the circumference of a circle to its diameter. The ancient Egyptians were acquainted with the fact that this ratio is the same for all circles, but Archimedes showed that it could be calculated by pure theory assuming only the qualitative axioms of Euclidean geometry. Similarly, according to Whittaker, Eddington assumed various *qualitative* laws of physics, e.g. the exclusion principle already mentioned, and deduced the *quantitative* laws, i.e. the exact values of the numbers known as the constants of nature.¹ Thirdly, as we have already stressed, the keystone of Eddington's theory is the assumption that the universe is finite.

Eddington's *uranoid* is a static Einstein universe of hydrogen² at zero temperature. Such a system, as we have seen, is unstable: if submitted to a slight disturbance it will either start to expand or contract. All the available evidence indicates that we should direct our attention to the possibility of expansion. In particular, we are led to this conclusion by comparing the calculated density of matter in the *uranoid* with the observed density of matter in the universe as deduced from counts of extra-galactic nebulae and current estimates of their masses.

According to Hubble, the average density of matter in the universe is not greater than 10^{-29} grammes per cubic centimetre.³ Eddington thought it unlikely that the total amount of unobserved matter in the world, such as cosmic dust and dark nebulae, could increase this density by a factor as great as ten. On the other hand, the average density of matter in the *uranoid*, given the mass and radius previously stated, is of the order of 10^{-27} grammes per cubic centimetre, i.e. about a hundred times greater than Hubble's figure and equivalent to about one proton per litre.

In his final memoir, submitted to the Royal Astronomical Society in 1944, Eddington pointed out that, if the *uranoid*

¹ For further discussion on this question, see Chapter IX.

² The hydrogen atom contains one proton and one electron. Recent work on the problem of the abundance of the elements indicates that hydrogen is probably the preponderant element throughout the actual universe.

³ It is now thought to be nearer 10^{-28} .

expanded until its radius had increased five-fold, its density would then be equal to that of the observed universe according to Hubble's estimate. Moreover, the observed rate of expansion (on the Doppler interpretation of the red-shifts) will agree with that calculated from Lemaître's theory to within the probable errors of observation. According to this theory, the limiting velocities of recession of the nebulae from each other after an infinite lapse of time will be about 572.4 kilometres per second *per megaparsec apart*. The figure given by Hubble and Humason for the present observed rate is 560, but this may be in error by as much as ten per cent. According to Lemaître's theory, the rate of recession when the radius is five times its initial value is about 540 kilometres per second per megaparsec.

Eddington assumed that expansion has a negligible effect on the 'constants of nature', which are determined by the finite static spherical uranoid. In his opinion, these fundamental constants are fixed for all time. This point of view is in striking contrast to that of Milne, who first directed his attention to theoretical cosmology in 1932. Dissatisfied with current theories he returned to first principles, regarding expansion as the fundamental feature of the universe. He pointed out that, in general, any system of uniformly moving particles, initially distributed at random in a finite region of space, would eventually be found to be receding from one another according to Hubble's relation, the fastest having travelled farthest. Compared with the actual universe, of course, such a system can only be regarded as a highly simplified model. If the nebulae are in fact all retreating from each other, it would appear that each is escaping from the gravitational field of the rest. With lapse of time their mutual gravitational attractions would steadily diminish, so that eventually the system would approximate to a set of freely moving particles exerting a negligible influence upon each other. Milne therefore thought it reasonable to assume that in this case the relative velocities would tend to become uniform. Thus, by neglecting gravitation, it should be possible to obtain some insight into the state towards which the universe is tending, and Hubble's law would appear to be its characteristic property.

If a swarm of particles expands uniformly from an initial

configuration of maximum congestion, the distance travelled by each particle during time t will be the product of t and its uniform velocity. With increasing lapse of time, this product will greatly exceed the particle's initial distance from any other member of the system, and hence will become an increasingly good approximation to the actual distance. Consequently, in this case, Hubble's law will arise in a form containing a number t which can be interpreted as the 'age' of the system, different configurations corresponding to different values of t . In the empirical form of Hubble's law, as stated in Chapter II, this 'constant' has a value of about 2,000 million years. It is convenient to denote this particular value by a special symbol t_0 . In Milne's cosmology t_0 plays a fundamental role, analogous to a new constant of nature.

Some three years before Milne developed this new line of thought, L. L. Whyte in a short but prophetic essay, *Archimedes or the Future of Physics*, pointed out that in each of the two great new physical theories of this century the fundamental role was played by a particular constant of nature: in Relativity by c , the velocity of light *in vacuo*, and in Quantum Theory by h , Planck's constant. He suggested that the next great advance in our understanding of nature would be associated with a new fundamental constant, and he prophesied that this would be concerned with the flow of time. "The idea that time may be an active factor in causation has the mathematical significance that 't' (for the system in question) must appear explicitly in the formulation of the law. . . . Such a law may claim to express the fact of historic, irreversible duration", or, as I prefer to call it, 'trend'.

Whyte showed how at the core of Newtonian physics lies the assumption that the elementary processes of nature are reversible, or would be if they could be isolated, and hence in this system of Natural Philosophy time does not appear as an explicit factor. "The question of the reversibility of natural processes," he wrote, "provides the key to a great intellectual struggle which is now in progress behind the complexities of philosophic and scientific thought. The issue can be formulated thus: Is there a real temporal process in nature? Is the passage of irreversible time a necessary element in any view

of the structure of nature? Or, alternatively, is the subjective experience of time a mere illusion of the mind which cannot be given objective expression? These are not metaphysical questions that can still be neglected with impunity. For just as Einstein made his advance by analysing conceptions such as simultaneity, which had been thought to be adequately understood for the purposes of experimental science, so the next development of physical theory will probably be made by carrying on the analysis of time from the point at which Einstein left it."

In the cosmological theories of Einstein, de Sitter and Lemaître new ideas were introduced concerning the character of universal space, but no corresponding advance was made in connection with the idea of time, except in so far as the phenomenon of expansion pointed to a finite rather than an infinite past. Milne, however, recognized that, as the phenomenon of expansion is irreversible, it must be intimately related to our awareness of the unidirectional flow of time.

As we have seen, the concept of physical time introduced into Natural Philosophy by Galileo was fundamentally geometrical in character. In geometry, the notion of order is based on the concept 'between'. This notion is reversible, for, if *B* is between *A* and *C*, then *B* is also between *C* and *A*. In our idea of subjective time-flow, however, the fundamental notion is of earlier and later. This notion is irreversible, for, if an observer regards the event *B* as later than the event *A*, it is impossible for him to consider the event *A* as later than the event *B*. The idea of time in classical physics was quite distinct from this subjective time of human experience. It was concerned only with 'duration' and not with 'trend'. The universe was regarded either as eternal or else as having been suddenly created by some external agent.

In classical physics, most of the fundamental laws of nature were concerned either with the stability of certain configurations of bodies, e.g. the solar system, or else with the conservation of certain properties of matter, e.g. mass, energy, angular momentum or spin. The outstanding exception was the famous Second Law of Thermodynamics, discovered by Clausius in 1850. This law, as usually stated, refers to an abstract concept called

entropy, which for any enclosed or thermally insulated system tends to increase continually with lapse of time. In practice, the most familiar example of this law occurs when two bodies are in contact: in general, heat tends to flow from the hotter body to the cooler. Thus, while the First Law of Thermodynamics, viz. the conservation of energy, is concerned only with time as mere duration, the Second Law involves the idea of trend. Milne developed his cosmology by taking this idea of trend to be fundamental, regarding the expansion of the universe as its supreme manifestation.

We have already seen that Milne showed that Hubble's law could be immediately explained by beginning with the idea of expansion if the counteracting effect of gravitational attraction were neglected. In 1934, however, partly in collaboration with McCrea, he showed that, if the effect of gravitation were taken into account, many of the familiar results of relativistic cosmology could be obtained using only classical ideas. In the last chapter we saw that difficulties arise if we imagine the universe to be an infinite static system in Euclidean space and subject to Newtonian gravitation. Once the possibility is admitted that the universe as a whole is not static, these objections do not necessarily apply. Milne and McCrea therefore examined the properties of an expanding universe subject to the laws of Newtonian physics. They discovered that the *local* properties of the relativistic models (with zero cosmical constant and zero pressure) in expanding spaces of positive, zero or negative curvatures are observationally the same as those of the respective Newtonian universes in which every particle has a velocity either less than the velocity of escape from the observer, or equal to, or greater than, this velocity. Hence they considered that the idea of non-Euclidean space was not essential to our understanding of the universe.

Encouraged by these successes, Milne proceeded to construct a new cosmological system from first principles. Whereas Einstein and his followers had found it necessary to introduce new concepts of space while retaining the classical concept of time, Milne, regarding expansion as fundamental, preferred to introduce a new theory of time while retaining as far as possible the classical concept of Euclidean space.

The new theory of Milne is called Kinematic Relativity, to distinguish it from the Special and General Relativity of Einstein and to emphasize that it is primarily based on a new theory of the measurement of time, space and motion. In Einstein's theory the classical concepts of the rigid measuring rod and the clock are both retained, although as Poincaré pointed out in 1906 distances are judged to be equal if light-signals traverse them in equal times. This criterion is an immediate consequence of the fundamental postulate that the velocity of light *in vacuo* is the same for all uniformly moving observers. Elaborating Poincaré's point, Milne maintained that the classical concept of the rigid rod was redundant. He, therefore, proposed to base his new theory of measurement solely on the observer's awareness of the flow of time and the method of signalling, e.g. by means of light or other electro-magnetic phenomena.

This policy was criticized by many authorities, including Eddington, as conflicting with the customary practice of laboratory physicists who regard length as the fundamental type of physical measurement. Nevertheless, the invention of radar has shown that Milne's ideas foreshadowed the most recent developments in metrology. In radar technique an estimate of the distance of a reflecting surface is provided by the time taken for a pulse of vibrations to return to the emitter, and the same principle was first suggested as the foundation of theoretical kinematics by Milne several years before radar was invented.

In 1933 the author showed that the Lorentz formulae correlating the space and time measurements assigned to the same event by two observers in uniform relative motion could be obtained by Milne's method without using any rigid body for measuring lengths. This was an advance on Einstein's procedure. Shortly afterwards, Milne discovered that his method could be applied even when the relative motion of the two observers was no longer assumed to be uniform.

The superiority of this method over previous methods of physical measurement is due to its dependence on fewer basic assumptions. Not only is it independent of the idea of the rigid body, but also of the traditional concept of a fixed scale of time.

It is, of course, assumed that the trend, or order, of events experienced by an observer is given, but no assumption is made regarding the rate at which they succeed each other. Since Newton it had been generally assumed that there is a unique scale of time in nature. Even after Einstein had discarded the idea of world-wide simultaneity, it was still thought that all natural phenomena suitable for use as clocks by a given observer must 'keep the same time'. For example, the daily rotation of the earth on its axis and its annual rotation about the sun provide two natural clocks which for most practical purposes 'keep the same time'. Milne and the author, therefore, broke new ground in investigating the possibility that there might exist natural phenomena, defining physically significant time-scales, which are non-uniform when referred to the rotating earth.

Two time-scales are said to be non-uniform, relative to each other, when events which appear to occur at equal intervals according to one appear to occur at unequal intervals according to the other. Two clocks which beat non-uniformly can, however, be correlated. A mathematical relation can be found between the times assigned by them respectively to a given sequence of events. This relation can be used to regrade, or convert, epochs on one scale to epochs on the other. As in the radar method, distances are measured by the time-intervals of reflected signals. Consequently, the distance, and hence the motion, of an object will be described by the observer in different ways according to his choice of time-scale. Indeed, it has been proved that, *in theory*, a particular scale can always be found so that a given object can be described as moving in any given manner either towards or from the observer. *In practice*, of course, this choice of scale must be restricted so that a *finite* epoch is assigned to any event which is actually experienced, i.e. 'lived through', by the observer. Also, as indicated above, the choice of scale must leave the *order* of these events quite unaffected.

These abstract ideas concerning the measurement of time and space have an immediate application to cosmology. Since the discovery by Roemer of the finite character of the velocity of light, it has gradually become evident that, as we peer deeper

and deeper into space, the world-as-seen at any instant is not the world-as-it-'is' at that instant, for the more distant objects are seen at an earlier stage in their history. To construct a map of the world at a given instant we must correct the world-picture for the effect of this time-lag. This correction will depend on our postulates concerning the scale of time. If we change from one scale to another, the world-map will change too. The possibility that more than one map of the universe may be physically significant suggests that some of the controversies in cosmology may be due to this cause. The situation is somewhat analogous to the mapping of the earth's surface. The Mercator projection, for example, is often more useful than a globe, despite the distortion which occurs as we approach the poles. In mapping the universe, Milne claims that there are two different time-scales and two associated world-maps, each of fundamental importance.

Following the hint suggested by his simple model of uniformly moving particles, Milne considered a system of such particles which all coincide at a particular instant.¹ This instant he regarded as the origin of time for the system. Strictly speaking, the system was assumed to exist only subsequently to this singular event, which he ultimately compared to the Creation. A unique particle can be found in this model moving in any given direction with any given speed less than c , the signal velocity. Milne assumed that the signal paths (light rays) in his model could be mapped as straight lines in Euclidean space. Thus at any epoch t , subsequent to the initial epoch zero, the system fills the *interior* of a Euclidean sphere of radius ct , the most distant particles appearing to recede with speeds arbitrarily close to c .

All the world-models which we have considered so far were based on the principle of homogeneity, each region having the same properties as any other region. Milne adopted a similar postulate for his model, which was constructed so that its appearance and history were the same according to each observer attached to a constituent particle. These observers were assumed to have similar clocks, by means of which each

¹ For a detailed account of the *a priori* foundations of Milne's model, see Chapter IX.

described the system as uniformly expanding. It is clear that the system must contain an infinite number of particles, each observer mapping the system as if it were bounded at distance ct at epoch t by an impenetrable material barrier of infinite density. For, if there were only a finite number of particles inside the Euclidean sphere of radius ct , there would exist a definite rim of particles, from each of which the appearance of the system could not be the same as from an interior particle. It is of the greatest significance that the kinematics of the Lorentz formulæ (See Chapter IV), unlike classical kinematics, do in fact permit *each* fundamental observer to describe the system as expanding equally in all directions from himself as centre. The *apparent* effect of infinite density at distance ct is due to the Fitzgerald-Lorentz 'contraction' factor, which tends to zero as the speed of recession approaches c .

In comparing this model with nature, Milne identified the constituent particles as nebulae. As in the models of relativistic cosmology, local irregularities of distribution were neglected, the model being regarded as a representation of the basic structure of the universe. For this reason, Milne called his system the 'substratum'. The time-scale corresponding to uniform expansion is denoted by the symbol t . At epoch t , the distance r of a particle moving with velocity v away from the observer is given by the relation, $r = vt$. Consequently, in this model, Hubble's law relating distance and velocity will be obeyed at all epochs. The present value of t is of the order of 2,000 million years.

Milne discovered another scale of time, τ , which can be used to construct a system of measurement by his 'radar' technique so that the relative distances of all the particles of the substratum are fixed. The geometry associated with this new scale of measurement is hyperbolic, light-rays being mapped as the analogues of Euclidean straight lines. The substratum fills the entire hyperbolic space, the total volume of which is infinite. Moreover, there is no origin of time on this scale, the total range of past time being infinite. The infinite past corresponds, of course, to the zero of time on the t -scale, which plays a similar role to the absolute zero of temperature: it lies outside all possible experience.

To show that these two scales of time are of physical significance when the substratum is taken as a model of the underlying structure of the universe of nebulae, we must associate them with definite natural time-keepers. Milne identifies the t -scale with the vibrations of a standard source of light of any given colour. Thus on this scale the time-intervals between successive crests of light-waves are uniform. Moreover, on this scale, the nebulae are receding with constant velocities. In our discussion of the Special Theory of Relativity, we explained that the apparent rate of a moving clock is slower than that of a similar clock kept by the observer. Hence the beats of the light-waves emitted by a standard source in a receding nebula should appear to be slower than those of the corresponding waves emitted by a similar source in the laboratory. Since the product of wave-length and frequency is always equal to the velocity of the light-wave, which is constant, it follows that an apparent slowing down of frequency will be associated with a corresponding increase of wave-length. Thus the light from a receding source should appear to be redder than the light from a similar static source in the laboratory. Consequently, the spectra of the nebulae are shifted towards the red.

Milne identifies τ -time with the standard clock of classical mechanics, to which the swinging pendulum and the rotating earth are usually sufficiently good approximations. He shows that, if time is measured on this scale, a particle moving freely in the field of the substratum and acted on by no forces, other than the 'pull' of the system as a whole, would obey Newton's first law of motion, i.e. it would move with uniform velocity in a straight line. Hence τ -time is the scale originally introduced into physics by Galileo and Newton. However, if standard light-pulses keep t -time, the intervals between successive beats will not be uniform in τ -time. There will be a gradual increase in the frequency, or number of beats in a unit interval of τ -time and a corresponding decrease in the wave-length. Consequently the light emitted by a standard atom in a distant nebula will appear to be redder than the light from a similar source in the laboratory because the light which we now receive from the nebula originated a long time ago when its natural wave-length was longer.

We see, therefore, that the red-shifts in the spectra of the nebulae can be explained in different ways, depending on how we choose to measure the flow of time. The crucial point in Milne's theory is the hypothesis that in nature some phenomena are uniform on one scale and some on another. For example, when the τ -scale is employed many of the fundamental constants of nature, e.g. the constant of gravitation and Planck's constant, are truly constant; whereas, when the t -scale is used, they are proportional to the time which has elapsed since the epoch zero. On the other hand, as we have just seen, the nebular red-shifts indicate that the frequencies of standard light-waves appear to vary with change of epoch when the τ -scale is adopted. Hence, whichever scale we choose, some of the fundamental quantities in physics must vary with lapse of time. Indeed, the red-shifts serve to remind us that we inhabit a changing world, unlike the static cosmos which has been the traditional background of physics.

Milne's work has an important bearing on the difficult problem of estimating the distances of the fainter nebulae. It will be remembered that these estimates are based on the assumption that statistically the nebulae can be regarded as of uniform absolute brightness so that their apparent brightness is a criterion of distance. The crucial problem, therefore, is the determination of the precise rule relating apparent brightness to distance. This rule will depend on the motion of the nebulae. Thus, if we regard the nebulae as stationary, we shall attribute greater distances to them than if we regard them as receding, because recession causes the amount of light received by us in a unit interval of time to be diminished.

The energy of a light quantum, or photon, is the product of its frequency and Planck's constant, \hbar . Since the apparent frequency of light radiated by a receding nebula is diminished, due to the Doppler effect, Hubble assumes that both the number of photons received in unit time and their energy are diminished, and so in his analysis of the observations recession contributes the same factor twice over to the reduction in apparent brightness. In Milne's theory, however, when the nebulae are regarded as receding uniformly, Planck's constant \hbar increases with lapse of time, thereby partially compensating for the

reduction in apparent frequency, so that, although the number of photons received in unit time is still diminished by the same factor as before, the energy of each is unaltered. Consequently, in Milne's analysis, recession has only half the effect on apparent brightness that it has in Hubble's investigations.¹

Thus, in Milne's theory, if the nebulae are regarded as receding, greater distances will be assigned to them than by Hubble. This result appears to be of considerable significance, because Hubble claims that if the nebulae were in motion then the data relating to the more distant nebulae would indicate that the universe was much more compact both in space and time than he believes possible. In particular, the age of the system would seem to be less than 1,000 million years. Consequently, Hubble rejects the possibility of an expanding universe, although he is quite unable to account for the red-shifts. Milne, however, claims that the universe can be regarded either as expanding or static. If it is regarded as expanding, it is not so congested in space or cramped in time as Hubble believes. Milne maintains that its age is approximately 2,000 million years.

¹ Nevertheless, it is sufficient to make the total apparent brightness of the whole infinite system finite.

CHAPTER VII

THE AGE OF THE UNIVERSE

ALTHOUGH there is at present no unanimity of opinion on the spatial extent of the universe, there is fairly general agreement that the total range of past time is finite. In Milne's theory, it is true that the apparent extent of past time is infinite according to the τ -scale. This scale, however, is an artificial scale introduced into physics for the sake of mathematical simplicity. We have seen that it can be identified with the time kept by the rotating earth, subject to certain corrections. Unless we consider events in the remote past or very distant future these corrections are small. However, we have reason to believe that there was a time when the earth, and indeed the whole solar system, was subject to very different conditions from those now prevailing. Consequently, the τ -clock must be regarded as a theoretical time-scale which agrees closely with the present rotation of the earth, but which diverges therefrom more and more as we extrapolate into the distant past or the remote future. We have no reason for assuming that there is any natural clock which keeps τ -time throughout the whole history of the universe. On the other hand the t -scale is identified with the vibrations of standard radiation, and Milne regards it as a more suitable scale for describing the full range of past time. As we have seen, according to the t -scale this range is approximately 2,000 million years. We will now consider other estimates of the age of the universe and compare them with this figure.

Until recent times all estimates of the age of the earth or the age of the universe were purely speculative. By a curious coincidence, just as the first theory of the structure of the world, according to Thales, concerned water, so the first theory of how to estimate the range of past time depended on the composition of the seas and oceans of the world. In the Middle Ages, Dante and others believed in the unity of all earthly waters. This concept was queried by the great Hohenstaufen Emperor, Frederick the Second, who has been called

the 'first modern man', mainly because of his scientific outlook which contrasted so sharply with the beliefs and prejudices of his age. Frederick was puzzled that sea-water was so salt, while river water was not. Nearly five hundred years later this distinction suggested to the astronomer Halley a possible method of estimating the age of the world. In a communication to the Royal Society in 1715 he pointed out that the sea had become salt because of the accumulation of saline material swept down by rivers. He regretted that the ancient Greeks had not "delivered down to us the degree of saltiness of the sea, as it was about two thousand years ago", so that the difference between the saltiness then and now could be used to estimate the age of the oceans.

Halley's suggestion was revived by Joly about fifty years ago. Assuming that the average annual amount of dissolved sodium removed by rivers from the land had remained constant throughout geological time, he estimated that about 100 million years would be required to provide the oceans with their present amount. It is now realized, however, that this method is much less accurate than Joly imagined. From the most recent investigations, Joly's assumption of past uniformity indicates an age of the order of 250 million years, but it can by no means be safely assumed that the rate of increase in the salinity of the oceans has been uniform. At present the land areas of the world are believed to be much more elevated than has usually been the case in the past, and rivers are much more active in consequence. Only a very rough estimate can be made of the effect of these and other influences, and the most that can be said is that the age of the oceans must be reckoned at least in hundreds, and perhaps in thousands, of millions of years.

Towards the end of the last century an alternative method of calculating the age of the earth was devised by Lord Kelvin. He pointed out that, as there is a measurable flow of heat through the earth's crust, indicated by the downward increase of temperature, the earth must be cooling and hence must have been hotter in the past. He calculated the epoch at which the earth was molten, finally concluding that this must have been between 20 and 40 million years ago. This estimate for the

age of the earth was in general agreement with his estimate for the age of the sun. Helmholtz had suggested that the sun maintains its enormous outpouring of radiation by continually shrinking and thereby releasing energy. If a body falls freely in a gravitational field it acquires kinetic energy, or energy of motion. If this motion is stopped, the energy thus acquired must be transformed in some way. In the case of a shrinking sun, Helmholtz suggested that the energy acquired as the outer regions fall in towards the centre appears as radiation. Kelvin calculated that the shrinkage of the sun to its present size could hardly have provided energy for more than about 50 million years of radiation.

Kelvin's calculations gave rise to a violent controversy between geologists and physicists. The geologists maintained that far longer periods of time were required to explain the known sequence of sedimentary strata and also to allow for the evolution of biological species. In particular, James Geikie showed in 1900 that the crustal compression resulting from 100 million years of cooling of the earth's crust would be confined to a shell which would be far too thin to accommodate the enormous thicknesses of folded rocks involved in the Alps and other great mountain ranges.

The flaw in Kelvin's method was not revealed until the turn of the century. Following the discovery of radium, Lord Rayleigh showed that this radioactive element occurs in common rocks all over the world. Thus the crustal rocks contain an unfailing source of heat of their own. Moreover, it is now known that there is a sufficient supply of radioactive elements in these rocks to make the net loss of heat extremely small, so that age estimates based on the rate of cooling are correspondingly increased. It is now thought that about ninety per cent. of the heat-flow through the rocks is due to radioactivity, and Kelvin's figure of 20 to 40 million years has to be multiplied about a hundredfold.

The decay of radioactive elements, however, has itself provided new and far superior methods for estimating the range of geological time. Early in the present century Rutherford and Soddy discovered that in each deposit of a radioactive element the number of atoms which disintegrate in unit

time is proportional to the total number of atoms of the element. The constant of proportionality was found to be independent of pressure, temperature and other physical conditions, and to depend only on the particular element considered. Thus the rate of decay of a given element could be used as a scale of time. When uranium and thorium decay they pass through a series of transformations; in particular, at one stage of uranium decay radium occurs. The ultimate stable end-products are helium and lead. By measuring the amount of these stable end-products in a rock containing uranium or thorium we can estimate the age of the rock. In practice more reliable results are obtained from studying the lead than the helium, as the latter is a gas and so more liable to escape. Fortunately, it is possible to distinguish between lead of radioactive origin and lead which is not produced in this way. If a radioactive mineral has not been affected by weathering and other changes, the amount of lead of radioactive origin now present in it depends on the amount of uranium and/or thorium now present and the period which has elapsed since the mineral originally crystallized. Hence, by measuring the amount of radiogenic lead present and comparing it with the amount of uranium and/or thorium, we can calculate the age of the mineral.

In practice this method can be checked by considering separately the three different isotopes, or varieties, of lead of radioactive origin and examining the consistency of our results. For example, the latest calculations agree to within a few per cent. in assigning about 255 million years since the close of the Devonian period, when the highest form of life is believed to have been fish. The Cambrian age, from which the earliest fossils date, probably occurred between 440 and 500 million years ago, while to the oldest mineral so far investigated, uraninite from Manitoba, has been assigned the tremendous age of 1,985 million years. Associated with this mineral are pebbles of still older granites and quartzites which must therefore be more than 2,000 million years old. Since the earth must be older still, this figure would seem to be a conservative minimum for the age of the earth.

Recently, Holmes has estimated the time that has elapsed since the earth's primeval lead was first modified by lead of

radiogenic origin. He used the data assembled by Nier concerning the relative abundance of the different isotopes of lead in samples of common lead minerals of different geological ages, and by extrapolation obtained a most probable estimate of 3,350 million years for the age of the earth. He claims that statistical analysis indicates that this result is unlikely to be seriously wrong.¹

It will be observed that this figure is somewhat greater than the figure of 2,000 million years assigned by Milne to the age of the universe when regarded as uniformly expanding on the t -scale. On the other hand, Eddington assigned an age of from ten to ninety thousand million years on the assumption that the universe is now expanding more rapidly than in the past. Nevertheless, the figure assigned by Milne to the age of the universe and that assigned by Holmes to the age of the earth are of the same order of magnitude. This agreement strongly suggests that the natural 'zero-time' of the universe as we know it may have occurred a few thousand million years ago.

Further evidence for this conclusion has come from the analysis of the helium and uranium content of meteorites. These are lumps of metal which enter the earth's atmosphere from outer space and succeed in penetrating to the earth's surface before being completely burnt up by their rapid passage through the air. It is believed that most, if not all, originate in the solar system. Paneth has subjected a number of meteorites to careful analysis and finds a considerable variation in their helium content, although the percentage of uranium and thorium is nearly identical in each. He concludes that this variation in the helium content may be due to a variation in age of the meteorites since solidification, the oldest being about 6,000 million years. Recently, however, it has been suggested that some of this helium may be of cosmic ray origin, in which case the age of the material has been overestimated.

Although the expanding universe hypothesis and the radioactive analysis of rocks and meteorites lead to estimates for the

¹ This claim is now disputed by Harold Jeffreys who published in November 1948 the results of an alternative statistical analysis of the same data. He suggests that Holmes may have overestimated the age of the earth by as much as a thousand million years.

age of the world of the same order of magnitude, until about ten years ago many arguments based on stellar observations seemed to show that our stellar system must be at least a thousand times older. This 'long' time-scale, of the order of a million million to ten million million years, was based primarily on the researches of Jeans.

One of the arguments for the long time-scale on which Jeans laid considerable stress concerned the observed approximation to 'equipartition of energy' among the stars. In 1911 Dr. Halm of the Cape Observatory correlated the average velocities of stars of various spectral types with their average weights or masses. He showed that for most types the product of their average weight and the square of their average velocity is the same. This product is proportional to the kinetic energy of the star. There is a famous law due to Maxwell which states that in a mixture of two or more gases there will be 'equipartition of energy', the heavier molecules moving more slowly than the lighter ones. Moreover, if the gases are suddenly mixed together in a vessel, it is possible to calculate the total time required for equipartition to occur. Halm's discovery suggested that the stellar system could be regarded as a gas in which the stars play the part of molecules. Given an initial random distribution of motion among the stars it should be possible to calculate the length of time needed for gravitational forces to speed up the slow stars and to slow down the fast stars until equipartition is observed. According to Jeans the required period is from 5 to 10 million million years.

This calculation was based on the assumption that the observed approach to equipartition was due to the effect of stars passing close to each other. In recent years reasons have been advanced for doubting whether, in fact, the observed trends towards equipartition are sufficient to indicate that stellar encounters have been causing them over an interval of the order of 10 million million years. For the trend towards equipartition should be universal, whereas there are important deviations. In particular, the kinetic energies of *B*-type stars are not more than half the mean kinetic energy of the majority of the stars. (The *B*-type stars are of a blue-white colour and their surface temperatures are from 15,000°C. to 20,000°C.,

compared with the $6,000^{\circ}$ of yellow-type stars like the sun.)

As Bok has recently emphasized, it should be kept in mind in all arguments concerning equipartition that the velocities of the stars must be determined with respect to a definite standard frame of reference. In practice we determine their apparent velocities with respect to the sun and then correct for the effect of the sun's local motion, i.e. its motion relative to the stars in its vicinity. During the last twenty years, however, it has been shown, notably by Oort, that our Galaxy is rotating. Consequently the velocities of the stars should be referred to the centre of galactic rotation, as the sun's local motion is less significant than its motion with respect to this centre. In particular, the *B*-type stars appear to move in very nearly circular orbits around the galactic centre.

In general, as Bok points out, "the whole problem of equipartition reduces itself to a basic problem of galactic dynamics and of stellar evolution, one that we have hardly begun to recognize as a problem and for which the solution is not yet in sight. Formulated concisely it reads: How can we explain certain observed correlations between the physical properties of the stars (such as mass, period, spectrum) and their dynamical, or orbital, characteristics in the galaxy? For the present we can only guess that from these correlations we may ultimately be able to draw important conclusions relative to the cosmic time-scale. We may, however, state with confidence that the equipartition argument can no longer be quoted in support of the long time-scale."

When we come to consider the detailed structure of the Galaxy, various objections to the long time-scale arise. If the stars have been in motion for more than a million million years, by now they should be distributed more at random than they appear to be. Eddington considered that on the long time-scale the Galaxy should by now be rotating uniformly; whereas, in fact it is not. Difficulties also arise in connection with the number of past rotations of our own and other galaxies. The period of rotation of our own Galaxy in the neighbourhood of the sun is about 250 million years, so that, according to the short time-scale, the number of past rotations is of the order of ten,

whereas on the long time-scale it would be of the order of ten thousand to a hundred thousand. Many authorities think that a nebula could not retain a spiral structure throughout so many rotations.

So far we have considered time-scale arguments relating to the earth, to the universe as a whole and to the stellar system or Galaxy. We must regard the age of the earth as providing a lower limit for the age of the universe, which can therefore be not less than 3,000 million years old, according to the radioactive or atomic time-scale. On the assumption that the extra-galactic nebulae are receding uniformly, the age of the universe would appear to be somewhat less than this. The discrepancy may be due to an overestimation of the age of the earth or to the fact that in the remote past the nebulae were moving more slowly so that the universe has taken a somewhat longer time to expand to its present state.

The assumption of uniform, or nearly uniform, expansion can be checked by testing whether the ages of celestial objects intermediate between the earth and the universe can also be shown to be of the same order of magnitude. Intermediate objects in descending order of magnitude include the Galaxy, or stellar system, star clusters and individual stars, such as the sun. We have seen that the original arguments favouring a long time-scale for the Galaxy are now regarded as much less cogent. It is therefore of considerable interest to determine the ages of star clusters and individual stars to see if they too are in general agreement with the short time-scale of a few thousand million years.

Our region of the Galaxy contains a number of loosely connected clusters of stars, for example the Hyades. The latter contains about a hundred and fifty stars within a distance of fifteen light-years from its centre, which is at about a hundred and thirty light-years from the sun. The motions of the individual stars differ by not more than about half a kilometre per second. In such systems each member star is prevented from wandering off into outer space by the gravitational pull of the whole system. There is, however, an important disruptive influence which tends to tear the cluster apart. This counter-acting force is the pull of the Galaxy as a whole, which is directed

towards the Galactic centre. In the case of loose clusters the interaction between individual members is relatively unimportant, and apart from the tug-of-war between the cluster as a whole and the Galaxy the most important forces influencing the development of the system are those due to the passage of other stars in its neighbourhood.

Bok has shown that clusters of this type fall into two classes according as their density is less than, or exceeds, a certain critical value depending on the distance of the cluster from the centre of the Galaxy and to a lesser extent on its shape. For spherical clusters, whose distance from this centre is comparable with the sun's, the critical density is of the order of three stars of the mass of the sun in a volume equal to that of a cube of side ten light-years. If the star-density is lower, then the cluster will be unstable and will disintegrate rapidly. If the star-density is greater, we can regard the cluster as stable, the rate of dissolution being much slower. The estimated density of the Hyades cluster is about two and a half times the critical density, and it has been calculated that its lifetime is of the order of 2,000 million years. During that period the probable number of encounters with other stars will have reduced the system to a state in which the Galactic pull can tear it apart.

The mechanism of disintegration of dense star clusters is quite different. In Galactic clusters like the Pleiades, the density of which exceeds the critical density by a factor of more than five, the pull of the Galaxy is no longer the dominant disruptive influence. Instead, the gradual impoverishment of these clusters is mainly due to the gravitational interaction between member stars. Associated with any gravitational system there is a definite velocity of escape. Knowing that a particular star in a cluster has a particular velocity, we can calculate the probability that this star will acquire the velocity of escape within a given interval of time. Stars acquiring this velocity will ultimately leave the system. The time required for this probability to amount to rather more than fifty per cent. has been defined by Chandrasekhar as a measure of the mean life of the cluster. The precise definition is, of course, a matter of convention. For the Pleiades cluster the mean life is of the order of 3,000 million years.

Although we do not know how clusters like the Pleiades

originated, it is thought unlikely that, with the general trend towards a purely random distribution of stars, there is much likelihood of new clusters being built up by chance encounters of individual stars. Hence these clusters, several hundreds of which are known to exist, appear to form a vanishing species. Most of them will cease to exist in a few thousand million years. We conclude that the rate of disruption of the Galactic clusters is convincing evidence in favour of the short time-scale for the evolution of the Galaxy.

The problem of determining the age of a typical individual star, such as the sun, is linked with the problem of stellar energy. Observation indicates that the sun is at present radiating energy at the rate of about 250 million tons a minute. Jeans suggested that the sun maintained this enormous torrent of radiation by the annihilation of matter, converting a fraction of its mass into radiant energy. On this basis he calculated that the sun could continue to shine for a billion years and lose less than ten per cent. of its mass. In the last ten years this suggestion has lost favour. An alternative method of energy generation is now generally accepted, and it has been shown that this is incompatible with the long time-scale.

Finally, further evidence for the short time-scale has been provided by observations on interstellar matter. Current estimates indicate that only about one-half of the mass of our Galaxy is to be found in the stars; at least an equal amount is scattered diffusely throughout interstellar space. For the determination of the cosmic time-scale, the most important feature is that this interstellar matter is not entirely gaseous as it contains dust particles. Lindblad was the first to consider the origin of these particles. He suggested that they were formed as condensations in the interstellar gas over a period comparable with the short cosmic time-scale. More recently ter Haar and van de Hulst have put forward the following theory. At fairly low densities molecules are formed from the interstellar atoms. The condensation nuclei, thus formed, will continue to grow through collisions, and particles with diameters of about a hundred-thousandth of a centimetre will be formed in about 30 million years, the slowest permissible rate producing the observed size in about 2,000 million years. We deduce

that the observed relative distribution of dust and gas points to the short time-scale, for on the long time-scale it is difficult to explain how so much of the gas has evaded capture by the solid interstellar matter.

In this rapid survey of the age of the earth, the stars, the Galaxy and the expanding system of nebulae, we see that calculations based on a wide range of diverse data are in general agreement, favouring the short time-scale. This situation is much more satisfactory than the position fifteen years ago, when the prevailing ideas on the age of the stars contrasted so sharply with the theory of the expanding universe. In 1932 de Sitter summarized the position as follows:

"The temptation is strong to identify the epoch of the beginning of the expansion with the 'beginning of the world', whatever that may mean. Now astronomically speaking this beginning of the expansion took place only yesterday, not much longer ago than the formation of the oldest rocks on the earth. According to all our modern views the evolution of a star, of a double star, or a star cluster, requires intervals of time which are enormously longer. The stars and the stellar system must be older than the universe!"

"What must be our attitude to this paradox? . . . I am afraid all we can do is to accept the paradox and try and accommodate ourselves to it, as we have done to so many paradoxes lately in modern physical theories."

The transition from this point of view to our present has an important bearing on our choice of a world-model. For in an attempt to resolve the time-scale paradox, de Sitter considered the possibility that the universe oscillates between a state of minimum and a state of maximum volume. In this way the present apparent expansion of the universe could be reconciled with the long time-scale. We now see that this not very attractive hypothesis is no longer required. Instead, all the available evidence points to the universe having expanded from an initial state of maximum concentration a few thousand million years ago.

THE STRUCTURE OF THE NEBULAE

OBSERVATION has so far failed to reveal any boundary of the system of nebulae, outside which there could exist similar systems. At present we have every reason for believing that the realm of the nebulae forms the ultimate background of the physical universe. Thus the nebulae themselves may be regarded as the individual bricks out of which the universe is constructed. So far we have concentrated attention mainly on the architecture of the building formed by these bricks. We will now consider some of the general characteristics and peculiarities of the bricks themselves.

Of the 100 million or so nebulae at present accessible to observation the vast majority can at best be recorded on the photographic plate as mere specks, barely distinguishable from the images of faint stars. A large number of nebulae appear somewhat brighter, but their images are still too small and faint for any details to be observed other than their elongation and the rate at which the brightness fades from the centre towards the edge of the image. Consequently, our knowledge of the structure of individual nebulae is mainly based on a few hundred exceptionally bright nebulae.

These nebulae have been arranged by Hubble in groups according to their characteristic features, and most of these groups have been found to form an ordered sequence which appears to be produced by the continuous variation of a fundamental pattern. The vast majority of the brightest nebulae are called 'regular', since they appear to be rotating more or less symmetrically about central condensations. The remaining few per cent. are called 'irregular', because they lack this feature. The regular nebulae fall naturally into two main groups, called 'elliptical' and 'spiral' respectively. About twenty per cent. of all nebulae so far classified are elliptical, and about seventy-five per cent. are spiral, but the true percentage of ellipticals may be considerably higher.

Elliptical nebulae range from globular objects through ellipsoidal figures to a limiting elongated form with a ratio of axes of about three to one. They are highly concentrated, but small patches of obscuring matter can occasionally be detected against the luminous background. The two main features employed so far for further classification are the shapes of the images and the luminosity gradients, or rates at which the apparent brightness falls off from the centre towards the edge. The former is the simpler criterion to adopt, as the shapes of the images and of the contours of equal luminosity can be rapidly determined. These shapes, however, are not true shapes, for they refer to the two-dimensional projected images on the photographic plate and not to the actual three-dimensional nebulae. Thus a circular image may refer to a highly elongated nebulae viewed in the direction of its axis of rotation and not necessarily to a spherical nebula. If we assume that the orientation of these axes is distributed at random, we can analyse the shapes of the projected images statistically, and thereby discover the relative abundance of the true forms. It appears that the true forms range from the globular to the lenticular, the latter being the more numerous.

There are two distinct types of spiral nebulae, normal and barred. In general, there are two spiral arms. In the normal spiral they emerge smoothly from opposite regions on the edge of a central condensation, which resembles a lenticular nebula. In the barred spiral the two arms emerge abruptly from the opposite ends of a bar of nebulosity stretching right across the central condensation. There appear to be at least twice as many normal spirals as barred spirals.

The normal spirals can be arranged in a definite sequence. Those at one end of the sequence exhibit a bright central core, the spiral arms being closely coiled. Progressing along the sequence, the arms increase at the expense of the central region and appear to unwind. At the end of the sequences the nucleus is relatively insignificant and the arms are widely open. Peripheral bands of obscuring material are frequently observed in normal spirals, particularly in those seen edge-on.

The barred spirals can also be arranged in a sequence beginning with a shape similar to the Greek letter θ . As the

sequence progresses the ring appears to break away from the bar at opposite ends so as to form a crude S, and the spiral arms gradually develop from the free ends. As in the development of the normal spiral, the arms grow at the expense of the central region, unwinding until finally a fully opened S-shape is observed.

The complete sequence of regular nebular types can be represented by a Y-shaped diagram. The stem is formed by the elliptical nebulae, with the globular systems at the base and the most elongated systems just below the fork. The two arms are formed by the normal and barred spirals respectively. Between the arms a few spirals of mixed type occur. The transition from the stem to the arm of barred spirals appears to be more or less continuous, but there seems to be a discontinuity between the stem and the arm of normal spirals. As Hubble remarks, "the suggestion of cataclysmic action at this point in the evolutional development of nebulae is rather pronounced."

Until recently, elliptical nebulae and the cores, but not the arms, of spiral nebulae appeared to be so concentrated that even with the 100-inch telescope it was not considered possible to resolve them into stars. It was therefore a major triumph of photographic skill when, in the autumn of 1943, Baade succeeded in resolving two of the nearest ellipticals and the central region of the great spiral in Andromeda. The outstanding significance of this feat, however, lay in the explanation of its difficulty. Baade pointed out that the former purely nebulous appearance of these objects was due to the abnormally low luminosity of their brightest constituent stars compared with the brightest stars in the arms of spirals. He suggested that we must henceforward recognize the existence of two types of stellar population in the nebulae. Type I, exemplified by stars in the sun's neighbourhood, is characterized by the most luminous stars with high surface temperatures (spectral types *O* and *B*) and open star-clusters. Type II is characterized by short-period Cepheids and stars such as are found in globular clusters. Whereas in elliptical nebulae, in the most closely coiled spirals and in the cores of other spirals only stars of type II are found, both types co-exist in the arms of open spirals, although type I usually predominates. Both types are found also in irregular nebulae.

Of the irregular nebulae the best known are our two nearest neighbours, called the Magellanic Clouds. They were first discovered by the Portuguese navigators of the fifteenth century when approaching the Cape of Good Hope, and were first described by a companion of Magellan on his voyage of circumnavigation. They are about 85,000 and 95,000 light-years away respectively, and their distance apart is about 35,000 light-years. Both clouds are highly resolved into stars, with no central condensations and no evidence of rotational symmetry. The main bodies of the clouds are roughly circular, their diameters being about 11,000 and 6,000 light-years respectively, the larger cloud being the nearer. This cloud is particularly rich in super-giant stars, one of which, S Doradus, is one of the most luminous stars known in the whole visible universe. It is a variable star of a peculiar type, its average brightness being about half a million times that of the sun. Recent evidence suggests that it may be a double star of two equal components.

The Magellanic Clouds belong to a typical small group of nebulae forming the so-called 'local group', of which our own Galaxy is a member. It includes all nebulae within about a million light-years from the sun, and contains at least thirteen members. The nearer nebulae in the general field lie at much greater distances (more than two and a half million light-years), and investigation of their detailed structure is much more difficult. The region occupied by the local group is roughly ellipsoidal in shape, with our own Galaxy near one end of the major axis and the famous nebula in Andromeda, M31, near the other. Unfortunately most members of the group are visible only from the southern hemisphere, where there are few large telescopes. Nevertheless, as Hubble has pointed out, the fact that our own Galaxy is a member of a group is a very fortunate circumstance,¹ for if it were otherwise our present knowledge of the nebulae would have required far more powerful telescopes than the 100-inch reflector.

The local group contains two triple nebulae: our own Galaxy and the Magellanic Clouds form one triple nebula, and

¹ It now appears that nearly all nebulae are members of groups, some of which contain only a few nebulae, others more, the greatest clusters containing several hundreds, of which most are elliptical.

the great nebula in Andromeda is a member of another set of three. This nebula is visible to the naked eye on clear moonless nights in autumn and winter as an elongated cloud about half the size of the full moon. It lies twenty-one degrees from the Milky Way circle. The nebula is a typical normal spiral with a relatively large nuclear region, which is the part visible to the unaided eye, and fainter arms, the outer parts of which are well resolved into stars. Its distance is still somewhat uncertain.¹ Despite numerous intensive investigations, based on hundreds of photographs of Cepheid variables, the distance can only be reckoned as three-quarters of a million light-years with an uncertainty of perhaps ten per cent. which may be partly due to our inadequate knowledge of the absorption of light by the dust in intervening space. The two neighbours of the Andromeda nebula are also at about the same distance.

The absolute brightness of M₃₁ appears to be roughly 1,700 million times that of the sun. If it were populated entirely by stars of the same type as the sun, knowledge of the absolute brightness would immediately inform us of the population. Many of the constituent stars are giants and super-giants with a much greater luminosity per unit mass than the sun. On the other hand, as in our own Galaxy, the vast majority of the constituent stars are probably much fainter per unit mass than the sun. In the present state of our knowledge, the luminosity method appears to inform us only that the mass of the Andromeda nebula is greater than 2,000 million times the mass of the sun, perhaps many times greater.

The main body of the Andromeda nebula, as shown in the best photographs with the most powerful telescopes, covers a surface area of the sky equal to about seven times that of the moon, and its length and width are 35,000 and 8,700 light-years respectively. If the elliptical form of the image is due to foreshortening owing to tilt, it is probable that we are viewing the system nearly edge on. Even with the aid of a telescope, however, the size of the nebula to the eye is much smaller than is revealed by good photographs. Using the micro-densitometer, a sensitive electrical apparatus for measuring photographs, it appears much larger still, the nucleus being surrounded by an

¹This uncertainty, of course, is not confined to M₃₁.

extensive haze of stars. Indeed, the area of the sky which we now know to be covered by the nebula has been increased ten times, and is approximately equal to seventy full moons. The major and minor axes of the image are roughly equal, and we conclude that the outer part of the system is more or less spherical. The whole system would appear to be like a flat wheel with a conspicuous hub, surrounded by a large spheroidal haze, probably composed of stars which are too faint to be resolved by our present methods.

M₃₂, the brighter and closer of the two satellites of M₃₁, is a typical elliptical nebula, the ratio of its axes being about four to five and its major diameter about 1,800 light-years. It is highly concentrated, and its brightness falls away rapidly from the core. It can be resolved into stars similar to those in the core of M₃₁. The other companion of M₃₁ is an abnormal elliptical nebula, NGC205, the nucleus of which resembles that of M₃₂, although it is considerably fainter.

Of the remaining seven members of the local group, two are classed as 'peculiar ellipticals', lying at rather more than 600,000 light-years from the sun, their diameters being about eight times those of the larger globular clusters of our own Galaxy. Two other systems, discovered about ten years ago, appear to consist of agglomerations of stars. These are the systems in Sculptor and Formax. The former is about 200,000 light-years distant, and the latter about twice as far. Their diameters are about 3,000 light-years and rather more than 6,000 light-years respectively. They differ from spirals in their lack of structural detail, and from ellipticals in their openness and transparency.

There is abundant evidence that most of the extra-galactic nebulae are permeated with a great deal of dust, and if during the course of time much of this dust gets swept up into the stars we should obtain comparatively transparent systems like these. There are no bright giant stars in these systems. As McVittie has remarked, "it would appear that agglomerations of stars of presumably small mass and luminosity can exist, ranging from loose systems such as those in Sculptor and Formax, through increasing degrees of concentration to the centre, until we reach a genuinely elliptical type such as M₃₂; that this type of

star population occurs in the core of a spiral such as M₃₁; but that to obtain giant blue stars we have to turn to the arms of spirals, where, apparently, these stars are alone found."

Of the three remaining galaxies of our group two are irregular nebulae similar to the Magellanic Clouds and of low luminosity. As Shapley has emphasized, it is rather disturbing to discover that so large a proportion of the nearest nebulae are confused and amorphous aggregations of stars, whereas we find in our surveys of the general field that only three or four per cent. are irregular. Unless we are driven to conclude that our part of the universe is not typical, we must query the completeness of our larger surveys. If the irregular nebulae in the local group were located 50 million light-years away, our best photographs would probably not reveal them. We conclude that there are likely to be many more small, faint and comparatively irregular nebulae than our surveys have so far yielded.

Finally, the most distant member of the local group is the open spiral M₃₃. It is a little more distant than M₃₁ and about 180,000 light-years away from it. The tilt is estimated at about thirty degrees, the ratio of the axes in the projected image being about two to three. The main body has a major diameter of about 12,000 light-years, visible in a good photograph with a large telescope. The micro-densitometer, however, reveals a diameter of nearly 20,000 light-years. It is somewhat brighter than the average nebula, the total luminosity being about 140 million times that of the sun.

The determination of the absolute luminosity of neighbouring nebulae, such as M₃₁ or M₃₃, is comparatively simple, once the distances of these objects are known. The estimation of their masses is much less straightforward. Using Newton's theory of gravitation, we can calculate the mass of a planet revolving round the sun, provided we know its distance from the sun and its period of rotation. Assuming that Newton's law can be applied to all rotating systems in the heavens, we can estimate the combined masses of double stars rotating about each other. The calculation of rotating nebular masses is more difficult.

The gravitational control of the stars in a typical spiral nebula differs from that of the solar system, where nearly all the mass is concentrated in the central sun, and the controlling

force therefore varies inversely as the square of the distance from the centre. It differs also from the gravitational field of a continuous discoidal system rotating as a cartwheel. Spectrographic studies indicate that the rotational motion of the outer parts is such that the actual gravitational field probably lies between these two extremes. Hubble originally estimated that the mass of the Andromeda nebula was 3,500 million times that of the sun, but we now know that he was observing only the central region. The central regions of nebulae are usually found to rotate like cartwheels, the speed of rotation at any point being proportional to the distance from the centre. In the outer regions the speed of rotation decreases as we recede from the centre. In the outer regions of M₃₁ the rotation period appears to be about 90 million years, while towards the rim of M₃₃ the rotation period appears to be about 200 million years. Various recent estimates for the mass of M₃₁ give about one hundred to two hundred thousand million times the mass of the sun.¹

Another method of estimating the order of magnitude of the mass of a typical nebula is to consider the individual motions of the members of a cluster of nebulae and calculate the gravitational pull which must be holding the cluster together so that it remains as a permanent structure. We can then calculate the mass of the whole cluster and hence deduce the average mass of the nebulae of which it is composed. Sinclair Smith has examined a cluster of thirty-two nebulae in Virgo, and has found an average mass of the order of 200,000 million suns.

Despite the general acceptance, during the last twenty or so years, of the theory that the great nebulae are similar structures to the Milky Way, one outstanding difficulty has not been fully resolved: the nebulae appear to be much smaller. Shapley's estimate of the dimensions of the Galaxy some thirty years ago was, however, an estimate of the dimensions of the enveloping system of globular clusters. There is now no doubt that the flattened wheel-like distribution of stars which constitutes the main body of the Milky Way represents the equatorial

¹ Is it now thought that over ninety per cent. of the total mass of our local group is concentrated in M₃₁, M₃₃ and our own Galaxy. It is significant that they are probably the only spirals in the group.

plane of the system of globular clusters, but there is no evidence that this disc reaches as far as the outer rim of the enveloping system.

In 1932 Hubble discovered on long-exposure photographs with the 100-inch telescope that the Andromeda nebula was also surrounded by at least 140 globular clusters. These extended far beyond the ordinary photographic limits of the spiral. Besides doubling the diameter of the system, this result strongly suggested by analogy that the limits of our own Galaxy fall short of the periphery of the region of its satellite globular clusters. Thus the difference in size between our own Galaxy and the Andromeda nebula is probably much less than formerly supposed.

To obtain more definite information concerning the extent of the main stellar system of the Milky Way, an elaborate search is being made at Harvard and in South Africa for remote Cepheid variables which are not members of globular clusters. Of course, the Galaxy has no definite rim; instead, the stars gradually peter out at great distances from the centre. In principle, we could define the edge of a galaxy as the region where the star-density in space falls below a definite fraction, e.g. one-thousandth, of the density at the centre. Unfortunately, this does not provide a practicable criterion. Instead, one has to make standard exposures with the same telescope and compare the images by a standard procedure. This method is satisfactory for the comparison of the extra-galactic nebulae with each other, but does not help in comparing them with the Milky Way. Indeed, the star-density in the region which we are disposed to regard as the outer rim of the Milky Way may be less than in the outermost detectable rim of the Andromeda nebula. Thus we may tend to exaggerate the dimensions of the former and underestimate those of the latter.

The main difficulty in investigating the Milky Way is due to the fact that we are located in just about the worst possible place for observation. Not only are we at the outskirts of the Galaxy, but we are also in such a dusty region that we cannot see very far in any direction in the Galactic plane. As Bok has remarked, "it is as though we were attempting, on a foggy day, to study the plan of a large city from the roof of a not-too-high

building somewhere in the suburbs". Because the layer of Galactic dust swells out to its greatest thickness in the direction of the centre, the Harvard survey is being made in the opposite direction, away from this hub. In this direction also there is a considerable amount of light-absorbing dust, and we suspect that our Galaxy is surrounded by a ring of obscuration such as we observe in so many of the external galaxies.

Nevertheless there are in this direction some comparatively transparent regions, through which we can faintly discern systems in outer space. It is in these regions that we can best seek to determine the confines of our own system, particularly by means of Cepheid variables. Of course, the distances of such stars cannot be determined accurately if we are unable to assess the amount of light absorption by the dust in intervening space. We can, however, estimate this effect by counting the number of external galaxies shown on long-exposure plates in the same direction. The effects of non-uniformity of distribution of the nebulae can only be circumvented by dealing with large areas involving many variable stars and tens of thousands of galaxies, thereby smoothing out irregularities and consequent errors.

According to our present knowledge, the distance of the anti-centre, or rim of the system in the opposite direction to the centre, is at least 15,000 light-years, but individual stars have been found at twice this distance. The distance of the centre, which is located in the direction of the great star cloud in Sagittarius, is from thirty to thirty-five thousand light-years. Until recently it was thought that the centre was hidden from us by a great dark cloud of obscuring matter, but in 1946 Baade suggested that parts of the Sagittarius cloud actually belong to the central nucleus. Thus we have an effective diameter of 100,000 light-years. Although we believe that the star-density decreases rapidly at this limit, there are scattered stars, including short-period Cepheids, extending to perhaps 15,000 light-years outside the boundary.

Similarly, the thickness of the stellar disc is not sharply defined. The stellar density decreases to one-tenth of that in the neighbourhood of the sun at a height of about 1,500 light-years, so that the vast majority of the stars lie in a layer, the thickness of which in the region of the sun is some 3,000

light-years. This layer increases considerably in thickness and density towards the Galactic centre. By analogy with external nebulae, it is probable that the spheroidal hub is 15,000 or more light-years in thickness. There are, however, scattered stars and globular clusters extending to more than 30,000 light-years above and below the central plane.

In the last few years much work has been done on the analysis of stellar distribution in the Galaxy. For example, it is found that, as we rise above the Galactic plane, the giant stars thin out much more rapidly than the dwarfs. Also, analysis of available star-counts indicates that in almost every direction the star-density drops with increasing distance from the sun. This diminution in the direction of the anti-centre was not unexpected, but the initial diminution in the direction of the Galactic centre was more surprising, particularly as there is a marked increase again beyond a distance of some two or three thousand light-years. By analogy with the extra-galactic nebulae, it would appear as if our sun were located in a minor spiral arm or knot of elongated shape. The initial decrease in star-density towards the Galactic centre would then be due to our passing beyond the spiral arm in which the sun is located, and the subsequent increase in star-density at greater distances would be due to our penetration of the main body of the Galaxy.

The globular clusters form a vast spheroidal super-system surrounding, and nearly concentric with, the highly flattened stellar system. This structure could be explained if the flat system were rotating about an axis, while the globular clusters were not. Study of the radial velocities of globular clusters reveals that they are all tending to move relative to the sun in a direction which is very nearly at right angles to the direction of the Galactic centre, with speeds ranging from two hundred to two hundred and fifty kilometres per second. This peculiar systematic asymmetry in their motion strongly suggests that the sun and its neighbours are all turning about the Galactic centre at a rate of more than 200 kilometres per second.

The earliest suggestion that the Galaxy is rotating was made by Sir John Herschel. The modern theory is due to Lindblad and Oort, although their work was partially anticipated by the Swedish astronomer Gyldén in 1871. If the Galaxy rotated like

a cartwheel it would be extremely difficult to detect this rotation, because relatively to each other the stars on the average would be stationary. Instead, the type of rotation for which we look is that of Saturn's rings. In a famous memoir, Clerk Maxwell showed nearly a hundred years ago that Saturn's rings would be unstable if they were solid, and that they must consist of a swarm of separate bodies. In a system of particles rotating about a centre of gravitational attraction the innermost particles will rotate more rapidly than the outermost, in order to counteract the stronger gravitational pull towards the centre.

In studying the problem of Galactic rotation we are almost entirely dependent on the determination of velocities in the line of sight, which can be measured spectroscopically by the Doppler effect. Only for the nearest stars can we determine motions transverse to the line of sight, by comparing their positions over a considerable interval of time. On the theory of Galactic rotation stars which are farther from the centre than the sun will tend, in general, to move more slowly, i.e. relative to the sun they will lag behind, whereas stars nearer to the centre will race ahead. It is this differential rotation which we must seek in our observations. There should be no average radial velocity in the line of the sun's motion nor in the line at right angles to this; for stars in the former line will have the velocity of the sun, and stars in the latter will be moving transversely to the line of sight.

The general differential effect can be visualized most easily by considering stars in a square centred on the sun and lying in the plane of the Galaxy, two sides being parallel to the direction of the Galactic centre. The stars on the outer edge will move backwards relative to the sun, while the stars on the inner edge will move forward. Consequently, the square will gradually be distorted into the shape of a diamond, two of its vertices being compressed towards the sun and two being drawn away from it. Hence along one diagonal the observed radial velocities should be towards the sun, and along the other they should be away from it. These diagonals should be equally inclined to the lines of zero radial velocity. Moreover, on the average, the radial velocities should increase with increasing distance from the sun. This effect was first traced by Oort in 1927, and he deduced that

the Galaxy was rotating about a centre in the direction of Sagittarius, thus providing an independent check on Shapley's estimation of the direction of the Galactic centre deduced from the direction of the globular clusters.

Oort's theory was a modification of that formulated by Lindblad in 1925-26. Lindblad assumed that the Galaxy was made up of a number of sub-systems, each in rotation, though at different rates, around the same axis perpendicular to the Galactic plane. He used this idea to explain the observed asymmetry in the motions of the so-called 'high-velocity' stars, i.e. stars moving with velocities of more than eighty kilometres per second. These stars all move towards one hemisphere of the sky. Lindblad suggested that, so far from being fast moving, these stars are laggards, and that their apparent high velocity is a reflection of the sun's velocity of rotation. We should, therefore, expect to see them all moving in the same direction, which should be exactly the reverse of the direction of the sun's orbital rotation. This effect has been fully confirmed by the investigations of Galactic rotation.

Lindblad's theory also threw new light on the phenomenon of star-streaming. If the stars were moving perfectly at random, we should expect to find as many stars moving relatively to the sun in one direction as in another. Kapteyn found, however, that there were two opposite points in the sky towards which more stars tended to move than in other directions. This line of preferential motion falls in the Milky Way and passes not far from the Galactic centre. Lindblad showed that, if the majority of stellar orbits were slightly elliptical rather than circular, this star-streaming effect would be reproduced. The theoretical significance of star-streaming was established by Eddington and Jeans, who showed that it is incompatible with strict dynamical equilibrium. In other words, it appears that our system has not yet reached a 'steady state'. This conclusion is reinforced by many of the peculiarities of the Galaxy, in particular by the existence of many moving groups of stars, for it is clear that differential rotation must have a shearing tendency on such groups, which after a few revolutions will tend to spread them out into a complete ring.

The theory of Galactic rotation implies that we can no

longer regard either the sun or the mean standard determined by the motion of the stars in our neighbourhood as providing an ultimate standard of rest in the universe. For these standards are local and are affected by an orbital velocity of some 250 kilometres per second. The centre of the Galaxy is a more fundamental standard, although it too can hardly be ultimate when we consider that, in all probability, the Milky Way has its own proper motion relative to the system of nebulae. The introduction of the Galactic centre as an origin of reference, however, has brought better order into the motions of the spiral nebulae. The rule that all these show velocities of recession was marred by a few outstanding exceptions among the nearest nebulae, e.g. the Andromeda nebula. We now realize that their apparent approach is partly the reflection of the sun's high speed of rotation in their direction.

Until recently it was customary to regard the Galaxy as consisting primarily of stars separated by vast distances compared with their diameters. For example, the distance of Sirius from the sun is nearly 50 million times the sun's diameter. The discovery of interstellar matter has, of course, greatly modified this picture, and we now regard the total mass of the Galaxy as more or less equally divided between the stars and the diffuse material lying between them. The discovery of interstellar gas dates from an important observation in 1905 by the German astronomer Hartmann. While studying the light emanating from the star Delta Orionis, he found that one of the lines in the spectrum, which was due to ionized calcium, behaved in a peculiar way. The other lines were broad and fuzzy, indicating that Delta Orionis actually consisted of two stars rotating about each other. The peculiar line of calcium, on the other hand, was sharp and distinct. In 1909 Slipher suggested that this line was due to calcium vapour lying between the star and ourselves.

The first systematic study of the subject was made by J. S. Plaskett. In 1924 he showed conclusively that the positions of certain lines of calcium and sodium in the spectra of a number of stars differed from their normal positions in such a way that it was clear that they could not arise in the atmospheres of these stars. He concluded that they must be produced by ionized

calcium and neutral sodium in interstellar space. Subsequent theoretical investigations by Eddington and Rosseland showed that calcium and sodium were the only reasonably abundant elements which could yield spectral lines observable by us under the conditions of stimulation prevailing in interstellar space. Eddington calculated that the density of the interstellar cloud was of the order¹ of 10^{-24} grammes per cubic centimetre, which is about one atom per cubic centimetre if, as in the stars, most of the atoms are hydrogen. As he remarked: "One atom per cubic centimetre does not amount to much. A portion of the cosmic cloud as large as the earth could, if compressed, be packed in a suitcase and easily carried with one hand."

In 1930 Plaskett and Pearce showed that the intensity of the interstellar lines is directly proportional to the distance of the stars in whose spectra they are observed, indicating that the interstellar cloud is distributed more or less uniformly throughout the intervening space, a result which has been considerably modified in recent years. They concluded that the cloud must extend to at least 6,000 light-years from the sun and probably to the confines of the system. They established, moreover, that the interstellar gas shares in the Galactic rotation, as it is nearly stationary with respect to the frame of reference determined by the average motion of the stars.

The reason why the interstellar cloud was not found earlier is because only certain stars show the effect discovered by Hartmann. First, the stars must be at least 1,000 light-years distant, so that the thickness of cloud in the intervening space is sufficient to produce a detectable effect. At the same time the star must be sufficiently bright to enable us to study its spectrum adequately. Consequently, the star must have an extremely high intrinsic brightness. Moreover, the star must not show in its own spectrum the calcium and sodium lines which appear in the spectra of interstellar vapour, for when these lines appear in the spectrum of a star's atmosphere they are usually so strong and broad that they could mask the sharp lines of the cosmic cloud. Hence, only comparatively few stars will show the required effect.

Further evidence for the existence of interstellar matter was

¹ It may be nearer to 10^{-25} .

provided by Trumpler's study of *similar* open Galactic star-clusters. Assuming that they all have the same absolute luminosity, Trumpler found that the diameters of the most distant were nearly twice as large as those of the nearest. As it is most improbable that this is a real effect, he concluded that the distances of the more remote clusters had been overestimated, thus causing him to overestimate their diameters. He attributed this to the effect of interstellar gas in diminishing the apparent brightness. He concluded that the absorbing matter was concentrated towards the Galactic plane in a layer six hundred to one thousand light-years thick.

Although the cosmic haze is roughly homogeneous, there exist regions in which it is much denser than the average. These include vast tracts a hundred or more light-years across; in general they are dark and opaque, or at best semi-transparent, but some are faintly luminous. These clouds are most numerous in the Milky Way, where they can be seen silhouetted against the background of more distant stars. Clouds of this type predominate in the direction of the Galactic centre. Strömgren has tentatively estimated that they occupy about five per cent. of space near the Galactic plane, and has calculated that their average density is probably more than a hundred times greater than that of the general haze. The average number of clouds encountered by light travelling in the Galactic plane in the neighbourhood of the sun is perhaps two in a thousand years. Many of these clouds contain small dust particles which scatter, or deflect, the bluer rays, thereby causing many stars to appear redder than normal, just as the light from the sun appears redder when seen through smoke or fog. This space-reddening of distant stars, particularly the highly luminous *O* and *B* stars, has been studied intensively in recent years, notably by Stebbins, and has become one of the most important sources of information concerning interstellar clouds.

In order to distinguish these clouds from the roughly uniform tenuous medium pervading the Galaxy, we call the latter the 'absorbing layer'. Hubble has shown how the existence of the absorbing layer accounts for the observed distribution of extra-galactic nebulae. The average numbers of nebulae per plate are greatest in the regions of the sky farthest from the

Milky Way, where obscuration from the absorbing layer is least. Towards the Galactic plane the numbers per plate decrease in a way which suggests that the obscuration is roughly proportional to the length of the light-paths through a uniform layer of obscuration. In comparing the theoretical world-models with the observed distribution of nebulae, the latter must, of course, first be corrected for the effects of this obscuration.

Recently, Oort has drawn attention to the important role which interstellar matter may play in the evolution of the nebulae. He points out that it is very prominent in the spiral nebulae, particularly in those seen on edge. He suggests that, if the dark matter in these is of comparable mass with the totality of constituent stars, it might cause instability leading to the formation of spiral arms, the origin of which has been the subject of speculation since their original discovery by Lord Rosse in 1845. Most attempted explanations have attributed the occurrence of spiral arms to some form of nebular instability. In recent years Lindblad has studied the stability of rotating stellar systems and the effects of a disturbance on an equatorial rim which is on the verge of instability. He claims that such a disturbance could cause spiral arms to be ejected from the nebula in the direction in which it is rotating. The observational evidence on this point is conflicting. Hubble maintains that the spiral arms trail behind the rotating nebula.

More than twenty years ago Jeans suggested that the occurrence of two spiral arms emerging from diametrically opposite points of a nebular nucleus might be due to tidal forces arising from the gravitational effects of the rest of the universe, but he was extremely dubious of the possibility of explaining the actual spiral character in terms of Newtonian gravitational theory. Writing in 1928, he summarized his views thus: "The only result that seems to emerge with some clearness is that the spiral arms are permanent features of the nebulae. . . . Their further interpretation forms one of the most puzzling, as well as disconcerting, problems of cosmogony. . . . Each failure to explain the spiral arms makes it more and more difficult to resist a suspicion that the spiral nebulae are the seat of types of forces entirely unknown to us. . . . The type of conjecture

which presents itself is that the centres of the nebulae are of the nature of singular points at which matter is poured into our universe from some other, and entirely extraneous, spatial dimension, so that, to a denizen of our universe, they appear as points at which matter is continually being created."

In 1946 a new theory of the spiral arms, somewhat akin to the speculation of Jeans, was developed by Milne with the aid of his kinematic studies in world-structure, in particular the theory of the two time-scales. In this investigation, however, instead of a 'singular point' at the centre of each nebula, there is a 'singular event' at which the whole system was created. This event may equally well be supposed to have occurred at the centre of any one nebula, as all coincided at the singular epoch, afterwards separating by their mutual recession. As rotation is the outstanding feature of motion within the nebulae, Milne first considered a family of coplanar orbits described by particles under the influence of a central nucleus. These orbits he regarded as circular when referred to the τ -scale of time and the scale of length associated with it. (It will be recalled that this is the scale of measurement implied by classical mechanics.) In transforming from the τ -scale to the t -scale, a fixed radius according to the former is described as a uniformly expanding radius according to the latter. Consequently, a closed circular orbit according to the τ -scale will become an open spiral orbit according to the t -scale. Milne considered a family of orbits described by particles which have passed through a fixed point (in t -measure) near the nucleus of a nebula, and showed that the present positions of such particles would lie along a spiral arm which describes a definite number of convolutions in the sense of the orbital motion, followed by the same number of convolutions in the reverse sense. He obtained a formula for this number in terms of the mass and radius of a nebular nucleus. For our Galaxy this yields a number which is approximately two.

Milne suggested that the doubling back of the spiral arms predicted by his theory might be the origin of the discrepancy between Hubble's and Lindblad's views concerning the sense of unwinding. He pointed out that there was no necessity for an actual spiral to show the full set of convolutions, both direct and

retrograde. According to his theory, ejection which occurred earlier than a certain critical epoch, roughly a thousand million years ago, will be represented only by the outer retrograde convolutions, and so if ejection ceased at some previous epoch no inner convolutions will be represented. Similarly, ejections since the critical epoch will be represented only by inner convolutions. He also showed how barred spirals might be those for which a substantial part of the ejection occurred near the critical epoch. Milne's theory of the possible doubling back of the spiral arms has received some observational confirmation by Lindblad. In two spirals he claims to have discovered faint outer arms which trail, in addition to the main inner arms which proceed in the sense of rotation. Lindblad also advanced an independent theory of this doubling back.

Promising as Milne's theory appears to be, it does not attempt to account for all the peculiar features of the spiral arms. It tells us nothing, for example, about their composition. In particular, we do not know why the bright blue stars appear to be restricted to these arms. Indeed, it is largely due to this phenomenon that the spirals are seen as such. Baade has photographed M33 in the infra-red, and finds that its spiral form is no longer reproduced. Consequently, if only small dim stars are considered, the spiral form must be much less pronounced. We now believe that the vast majority of stars cannot be bright stars, for the total luminosity of a nebula would be vastly greater than observed if a considerable fraction of its mass were in the form of such stars. Hence, the appearance of a spiral shape may be a luminosity rather than a mass distribution effect. Furthermore, according to the latest theories of the source of stellar energy, it appears that the lifetime of a bright blue star is only some hundred million years, and hence a complete theory of the spiral arms should account for the origin and maintenance of these comparatively shortlived stars.

Hubble's classification of the nebulae according to their shape and degree of compactness is strongly suggestive of an evolutionary sequence, although the direction in which evolution proceeds along this sequence is still uncertain. The classical theory of nebular evolution originated in the speculations of Kant and Laplace. They considered a vast spherical homo-

geneous gaseous mass in a state of rotation. As it cools this mass contracts and increases its speed of rotation. It gradually flattens and assumes a lenticular shape. Instability at the rim causes a ring of material to part company from the main system. As the latter contracts still further, additional rings are formed. This hypothesis was originally advanced to account for the origin of the solar system, but it has long been realized that it is inadequate for this purpose. Instead, Jeans has suggested that Laplace's process takes place on a vastly greater scale than he imagined. The primitive nebula is not a single sun in the making; but it contains substance sufficient to form hundreds of millions of suns. In the last few years Hubble has emphasized the ring-like character of the outer regions of many of the extra-galactic nebulae, and it *may* happen that in the majority of the nebulae the ejection of matter from the nucleus takes place not merely at one or two points but along the whole periphery.

All theories of the origin of the stars and nebulae are still mainly speculative. We have already seen that, if all the matter in the observable region of the universe were uniformly scattered throughout space, the density would be at least 10^{-29} , perhaps 10^{-28} , grammes per cubic centimetre.¹ This is almost inconceivably low. As Jeans has remarked, "If the amount of air which occupies the space of a pinhead in our atmosphere were reduced to this density, it would occupy a hundred million cubic miles." According to Eddington, the original density of the universe when it began to expand may have been a hundred times as great as it is now. Jeans calculated that in gas of this density condensations would have to be reckoned in terms not merely of the sun's mass but of thousands of millions or hundred of thousands of millions of times the sun's mass in order to persist, depending on the average velocity of the particles composing the gas. Thus the original condensations would be comparable with nebulae rather than with stars.

According to Milne's theory of world-structure, the original density of the universe must have been vastly greater than it is now. In this case, we should expect the primeval universe to have been in a state of enormous compression and tremendous

¹ It may even be as high as 10^{-27} grammes per cubic centimetre, if an overwhelming proportion of matter is non-luminous.

temperature. Indeed, the universe may have begun as highly concentrated neutron or even radiation. (In the beginning God said, 'Let there be light'.) Lemaître has suggested that the universe began as an enormous radioactive super-atom, which afterwards passed slowly through the unstable Einstein configuration while the cosmical repulsion overcame world-gravitation. These more startling speculations on the origin of the universe have the advantage of providing the peculiar physical conditions which appear to be required for the synthesis of the heavy radioactive elements out of the lighter atoms, such as hydrogen, although Hoyle has suggested that they may originate in the explosions of supernovae stars.

On all questions relating to the most distant past and ultimate future, theory is still *highly* speculative, but there is one important problem of this kind on which Milne's theory seems to mark a definite advance. The origin of rotation in the universe has long been a mystery. The earth spins about the sun, the sun spins about the Galactic centre, the nebulae appear to spin about their axes of symmetry. According to classical physics the total amount of 'spin' (angular momentum) in the universe is conserved for all time. In this case its origin would appear to be a complete mystery. But, according to Milne's theory, using the *t*-scale of time, a very small amount of angular momentum in the 'high and far-off times' would in the course of ages gradually increase indefinitely. It is therefore not surprising that we see the spiral nebulae spinning like gigantic Catherine-wheels.¹

¹ Since this chapter was written, it has been announced from Mount Wilson that as far as mass is concerned the main substance of the great nebula in Andromeda appears to consist of a non-spiral distribution of stars of Population II, the spiral structure formed by stars of Population I being embedded in it and playing the role of an incidental feature.

CHAPTER IX

COSMOLOGY AND THE *A PRIORI*

... Or if they list to try
Conjecture, he his Fabric of the Heav'ns
Hath left to their dispute, perhaps to move
His laughter at their quaint opinions wide
Hereafter, when they come to model Heav'n
And calculate the Starrs . . .

MILTON, *Paradise Lost*, viii, 75-80.

AT the beginning of this essay, attention was drawn to a remark by Pascal on the incompleteness of our observational knowledge of the universe: "But if our view be arrested there let our imagination pass beyond." More recently, de Sitter has reminded us that "the universe is a hypothesis". From the physical point of view, everything outside the observable region is pure extrapolation. Owing to the limitations of our empirical knowledge of the universe, cosmology appears to be more speculative and less securely founded than most other branches of physical science. As these developed they separated from their common parent, philosophy; but, despite the remarkable advances of the last thirty years, cosmology is still regarded by many as a form of philosophical speculation. Although we have travelled a long way since the pre-Socratic philosophers first attempted to discover the essential nature of the world by pure thought, *a priori* arguments still play a leading role in cosmological enquiry. Now that the *a priori* has been discredited by so many philosophers and men of science, cosmology would seem to be its last important refuge. The great intellectual syntheses of the famous metaphysical philosophers have crumbled into ruins. Will not the same fate overtake the theories of the twentieth-century creators of the *a priori* world-models?

There is, however, a significant difference between the arguments of the metaphysical philosophers of the past and those of the theoretical cosmologists of today. The former were

primarily concerned with the ontological *a priori*, i.e. with being, the latter with the epistemological *a priori*, i.e. with deductive knowledge. A famous example of a philosophical argument concerning being is Descartes' *Cogito ergo sum*. Philosophies of being were propounded by the great Continental rationalists of the seventeenth century, in particular by Spinoza who said that being is one, and by Leibniz who said that being is a characteristic of a plurality of monads united by a 'pre-established harmony'.

On the other hand, the great British empiricists of the late seventeenth and early eighteenth century, Locke, Berkeley and Hume, concentrated on the problem of knowledge. Locke rejected the theory that our minds can entertain any ideas prior to experience of the external world. All our knowledge comes from the impression of external events on our minds, which are purely receptive. Berkeley attempted to identify being and knowledge. For him 'to exist' meant 'to be apprehended by a mind'. *Esse est percipi*: to be is to be perceived. Berkeley having demolished matter, Hume attempted to demolish mind. Thus philosophical enquiry into the nature of knowledge appeared to lead to complete scepticism, until Kant asked his epoch-making question: "How is knowledge possible?"

Kant showed that in studying nature we are both actors and spectators, for knowledge is shaped by the intellect as well as by the external world. Bergson has placed Kant's theory of knowledge in its historical perspective in the following passage: "Spinoza and Leibniz had, following Aristotle, hypos-
tatized in God the unity of knowledge. . . . For the ancients, science applied to *concepts*, that is to say to kinds of *things*. In compressing all concepts into one, they therefore necessarily arrived at a *being*, which we may call Thought, but which was rather thought-object than thought-subject. . . . God was the synthesis of all concepts, the idea of ideas. But modern science turns on laws, that is, on relations. Now a relation is a bond established by a mind between two or more terms. A relation is nothing outside the intellect that relates. The universe, therefore, can only be a system of laws if phenomena have passed beforehand through the filter of an intellect. Of course,

this intellect might be that of a being infinitely superior to man who would found the materiality of things at the same time that he bound them together: such was the hypothesis of Leibniz and Spinoza. But it is not necessary to go so far, and, for the effect we have to obtain, the human intellect is enough: such is precisely the Kantian solution. . . .”

In their most profound formulation, the cosmological theories of Eddington and Milne are both developed from *a priori* epistemological principles. In these theories the universe is regarded as the background of natural phenomena, and the respective epistemological premises, from which each author claims to deduce fundamental laws of nature, are statements about scientific method. In other words, they are not statements about the detailed structure of the world, but about how we propose to analyse this structure. As Eddington has said, “We have to show not that there are N particles in the universe, but that anyone who accepts certain elementary principles of measurement must, if he is consistent, think that there are.”

As we have already indicated, Eddington, in his *Fundamental Theory*, claimed to have established many of the quantitative laws of physics as consequences of certain qualitative principles of scientific method. He argued that the idea of measurement implies a certain structural pattern of thought which, when applied to the analysis of natural forces, automatically produces the quantitative laws of gravitation, electromagnetism and so forth. He maintained that any measurement involves four entities, two for the observable to be measured and two for the standard with which it is compared. Thus, when we measure a length, we associate, at least in principle, the two end-points of the length with two marks on a ruler, and Eddington claimed that all physical measurements can ultimately be reduced to ‘pointer readings’ of this type. The most primitive entity of thought has only one of two possible properties: existence or non-existence. Consequently the most primitive measurable is determined by four entities (end-points), whose existence attributes are independent. It will, of course, exist only if all four entities exist. The two possibilities of existence or non-existence are called the ‘eigen-values’ of

an entity. A measurable, therefore, has sixteen such 'eigen-values', since sixteen is the product of four twos; only one of these sixteen 'eigen-values' will correspond to the existence of the measurable.

Constructing his theory on this concept, Eddington showed how the principles of physical measurement can be made to depend on a peculiar type of algebra (a 'linear associative algebra'), in which there are sixteen fundamental 'numerical operators'. He maintained that this algebra provides the structural pattern of the fundamental laws of physics. In particular, Eddington believed that it accounted for the fact that we assign three dimensions to space and one to time, and that all electrons carry the same charge, a fundamental principle of nature which has not been explained by any other theory.

Of course Eddington did not claim that it was possible to deduce *a priori* the complete structure of the physical universe, the sun, the moon, and all the stars, from the rules of measurement. Instead, he maintained that the fundamental numerical constants occurring in the general laws by which we select and order natural phenomena, e.g. the ratio of the masses of the proton and electron, the fine structure constant in the analysis of atomic spectra, etc., can be calculated without reference to particular measurements made in the laboratory. In other words, these universal numbers need not be obtained as inductive generalizations from a limited number of measurements of particular phenomena; instead, they arise as deductive consequences of the general principles of metrology and the theory of knowledge.

The vital stage in any *a priori* theory of physical phenomena, such as Eddington's, is the identification of the various symbols of the abstract structural pattern with the concepts employed in observation and experiment. Eddington compared scientific method to a fishing net, pointing out that many of the characteristics of the fish which can be caught by the net could be predicted by studying its mesh. Those characteristics which could not be predicted in this way he regarded as 'irrelevant', at least for physics. Hence, the properties which are considered significant for physics ought to be completely specified by his method.

That this method is incomplete in this respect appears probable when we consider the apparently self-contradictory situation that another *a priori* world-model and system of natural philosophy has been constructed by Milne. Despite mutual inconsistencies, the two are not necessarily mutually exclusive, for they may provide complementary perspectives of the physical universe. And in a general way we can dimly see how this might be; in Eddington's theory the spatial aspect is more thoroughly analysed than the temporal, whereas in Milne's, primary emphasis is laid on the flux of time. Eddington's concept of physical measurement is associated with the idea of the ruler, Milne's with the idea of the clock.

There is an interesting parallelism between the ways in which Eddington and Milne developed their respective theories of the universe and its laws. First, each began with a definite world-model and analysed its properties in detail. In Eddington's case this was the Einstein universe, in Milne's the uniformly expanding universe suggested by the Special Theory of Relativity. Despite many novel features, this phase of their respective investigations was in general agreement with the methods adopted by other theoretical cosmologists.

In the later phase of their work, the break with traditional methods was radical. Each attempted to deduce the model he had originally accepted as given. Eddington, as we have seen, claimed to have deduced his system from principles of physical measurement suggested by the technique of the ruler, and elaborated a new philosophy of physics to explain and justify his procedure. Milne, however, while devising an alternative theory of measurement, in which the clock displaced the ruler as the fundamental concept, deduced his system from a theory of natural philosophy which has been shown by the author to have much in common with the ideas of Norman Campbell. Natural science was described by Campbell in a treatise, *Physics: the Elements*, published in 1920, as the study of those judgments concerning which 'universal' agreement can be obtained, at least in principle. On this view, scientific method is the interpretation of phenomena by a principle of 'uniformity' and 'communicability'. Consequently, a community of (hypothetical) observers is a prerequisite of natural philosophy,

and the equality of status of such observers becomes the root of the relativity concept.

In the *a priori* development of Kinematic Relativity, Milne and the author began with the abstract concept of an observer as the kinematic analogue of the point in geometry and the massive particle in dynamics. The observer is assumed to have position and to experience a temporal before-and-after sequence of events. The essential feature of this sequence is its irreversibility. In geometry any two points determine a line, their join. In kinematics, any two observers determine what we call a 'signalling process'. Each is related to the other by a communication system, which is exemplified in nature by light. The equality of status of observers is defined by a postulate concerning their measurement of time-flow and the signals which pass between them. The sense of time-flow is given, but its measurement is arbitrary. Any correlation between the system of numbers in order of increasing magnitude and the totality of (hypothetical) events in the observer's immediate experience defines a possible clock. Consequently, whereas General Relativity begins geometrically, Kinematic Relativity begins arithmetically, and its objective can be described as the 'arithmetization of physics'.

Two observers are said to be equivalent if their clocks are chosen in such a way that when signals are emitted by either at the same time they will be received at equal times by the other. The distances between pairs of observers can be determined by the radar technique described in Chapter VI. A continuous distribution of hypothetical observers, each pair of which are equivalent, is called an 'equivalence'. This concept has a remarkable property; each observer can graduate his clock so that when using the radar technique he can describe every other observer as receding with uniform velocity, the whole system expanding from a point source at a particular epoch. Thus an equivalence is a structure of the type originally devised by Milne as a model of the expanding universe of nebulae, but in this deeper analysis uniformity of expansion is no longer *postulated*. Instead, it arises as a mode of description of an abstract kinematic framework, i.e. of a system of observers with congruent systems of measurement.

Milne maintains that an equivalence is a unique concept; that is to say, all equivalences are conformal, any one equivalence being convertible into any other by an appropriate regraduation of time-scale. According to the choice of time-scale, the motion pattern of the equivalence varies. In particular, clocks can be chosen so that the equivalence is described as motionless or static. By concentrating attention on this zero-motion pattern, we first discover that, corresponding to the various logically possible types of geometry, there are different logically possible types of equivalence. Consequently, the kinematic analysis of Milne provides the foundation of natural geometry. In applying geometry to nature, we usually begin by postulating, at least implicitly, a static background to the ceaseless flow of events. In Kinematic Relativity, however, this background is automatically constructed by the theory.

To isolate a particular geometry, and hence a unique equivalence, some further postulate is required. Milne has devoted attention to a particular geometry, but his reasons for so doing have not been regarded as conclusive by his critics. For example, Walker claims that at least three different equivalences have equal epistemological status: those for which, in the motionless pattern, geometry is either Euclidean, spherical or hyperbolic. Milne's case is the last.

What *a priori* epistemological arguments should guide us in choosing a limited number of geometries out of the innumerable logically possible types? We can be guided by our concept of scientific method with its emphasis on uniformity, as we require our geometry to provide a suitable background for describing natural phenomena in accordance with this concept. Thus, the background should itself be 'uniform'. A similar point was made by Whitehead in developing an alternative theory to Einstein's General Relativity some twenty-five years ago. "As the result of a consideration of the character of knowledge in general, and of our knowledge of nature in particular, I deduce that our experience requires and exhibits a basis of uniformity, and that in the case of nature this basis exhibits itself as the uniformity of spatio-temporal relations. . . . The structure is uniform because of the necessity for knowledge that there be a system of uniform relatedness, in terms of

which the contingent relations of natural facts can be expressed. Otherwise we can know nothing until we know everything."¹

The problem of characterizing uniform geometries was first studied in 1868 by the great German physiologist and physicist Helmholtz. He was led to it by his researches in physiological optics, in particular by those bearing on the localization of objects in the field of vision. He examined the spaces in which the properties of rigid bodies are not affected by translation and rotation. Such spaces will provide a suitable background for a system of congruent measurements by equivalent observers. As a result of Helmholtz's pioneer investigations and subsequent researches, notably by the Norwegian mathematician Sophus Lie and (with particular reference to Milne's theory²) by Walker, it appears that there are three geometries which exhibit the required uniformity: Euclidean, spherical and hyperbolic.

So far, an equivalence has been regarded as an abstract kinematic pattern. However, in dynamics attention is directed to the problem of determining whether a given material system maintains its configuration. In Milne's theory, this transition from kinematics to dynamics is made by associating with each observer a 'massive particle' defined by a 'causal law'. In a penetrating analysis of the axioms of dynamics, the French mathematician Painlevé pointed out that the aim of natural philosophy has been to deduce the phenomena of motion rigorously from the principle of causality. In the past this principle has usually been taken to imply that, when the same conditions are realized at two different instants in two different parts of space, the same phenomena reproduce themselves, only transported in space and time. Painlevé showed that classical physics was based on the assumption that at every point in space and at every instant in time a measure of length and a scale of time can be chosen so that this principle of causality is satisfied.

In Kinematic Relativity, on the other hand, there is no need to invoke this postulate with its arbitrary elimination of the

¹ Cf. the criticism already made of Einstein's space-time in General Relativity. The non-uniform geometry of space-time is assumed to depend on the distribution of matter, but the latter cannot be known until we know the geometry of space-time.

² In which no appeal is made to the properties of rigid bodies.

possibility of variation in time. Instead, the equivalence of massive particles, or 'substratum', needs to be submitted only to a principle of 'identity' and the principle of 'sufficient reason'. The former asserts that each particle should be anchored to a definite observer, while the latter implies that the symmetry of 'causes' must be reflected in the symmetry of 'effects'. Consequently, since each particle is at a centre of symmetry of the motion-pattern of the equivalence, it follows that each massive particle must be at a centre of symmetry of the mass-distribution; otherwise there might be a tendency for each particle to break loose from the observer to whom it is initially attached. Hence it can be shown that each observer will describe the substratum, its laws and properties, in the same way.

The substratum, being a continuous distribution of matter, is similar to a liquid. Another system with similar properties of homogeneity and symmetry about each observer can be constructed on the analogy of a gas, by surrounding each fundamental particle of the substratum with a haze of subsidiary particles moving with all speeds up to that of light. Such a system was analysed in great detail by Milne in the hope that it might enable us to reproduce many properties of the universe which are not revealed by the substratum, but it is now apparent that one of its main functions is to enable us to deduce certain properties of the substratum, by regarding the latter as the limiting case when the 'gas' is diluted indefinitely. By considering the substratum in this way Milne showed that the motion of any particle projected therein, subject only to the gravitational pull of the whole system, can be calculated without further postulates. When the substratum is described as static, the motion of any such particle is uniform. This result is reminiscent of Newton's law of inertia, but, whereas Newton's law referred to the motion in empty space of a free particle, the new law describes the motion of such a particle in the field of a homogeneous massive system which, as we saw in Chapter VI, can be compared with the statistically smoothed-out universe of nebulae.

We are now in a position to review the classical difficulties concerning absolute space and rotation raised in Chapter III.

Referred to empty space, the law of inertia led to the view that space is relative and that no meaning can be attached to absolute motion. On the other hand, the properties of rotating bodies indicated that space must be absolute, for a distinction can be made between a body which is rotating and one which is not. (For example, the flattening of the earth at the poles and the motion of Foucault's pendulum indicate that the earth turns on its axis.) A similar difficulty also arises when we consider two bodies or frames in relative acceleration. In classical physics absolute acceleration can be measured through the agency of force, and yet in empty space there should be no reason why of two frames in relative acceleration one should be regarded as an inertial frame and the other as absolutely accelerated. Moreover, in classical physics, the property of inertia has to be accepted as a *fiat* of nature.

According to the new view, all these difficulties vanish. Empty space is replaced by the 'smoothed-out univers.'. Relative to this we can say that one body moves uniformly while another is accelerated; one frame is rotating while another is not; and so forth. Inertia is due to the pull of the whole material universe, when all local irregularities have been smoothed out. Deviations from uniform motion are therefore due to the irregularities of distribution in the actual universe.

Milne has investigated the hierarchy of laws of different degrees of complexity which flow from the defining characteristics of the substratum, just as, for example, in Euclidean geometry we study the hierarchy of curves, etc., to which the axioms give rise. In this way, with the aid of a limited number of special postulates of a general character, e.g. that space has three dimensions, he has deduced laws which are formally similar to the classical laws of gravitation, electromagnetism, etc. In particular, he has derived an important relation expressing ρ , the average density of matter in the 'universe', in terms of the age of the 'universe' (in t -time), and G , the 'present value' of the 'constant' of gravitation, which in the t -measure varies according to the age of the 'universe'. This formula,

$$G \rho t^2 = \frac{3}{4\pi},$$

yields a theoretical value for the average density of about 10^{-27} grammes per cubic centimetre.

Despite their divergent points of view, Milne and Eddington agree in their Kantian emphasis on the function of the mind in arranging the maze of natural events according to the mind's own canons of order. In their view, the essential role of the *a priori* is contained in the principle of uniformity, regarded as an epistemological policy rather than as a metaphysical creed. Both Milne and Eddington have been widely criticized. For example, the philosopher A. J. Ayer has attacked Milne's claim that the laws of nature can be formulated in such a way that physics has the same deductive character as pure geometry. He writes: ". . . the question whether physics can be made to attain the status of geometry has nothing directly to do with the character of the physical world or even with the character of our knowledge of it. It is simply a matter of one's being able to organize the accepted law of physics into a self-consistent deductive system, and then choosing to regard the premises of this system, not as propositions about matters of fact, but as implicit definitions. . . . For no one would say that a proposition expressed a law of nature merely because it was assigned a place in some self-consistent abstract system." Milne's theory, however, like Eddington's, is not just 'some self-consistent abstract system'. It claims to be the product of a definite concept of scientific method, formulated in accordance with the essential spirit in which physics has been developed in the past.

The principle of the uniformity of nature originated with Thales. In his cosmology we first encounter the idea of the underlying unity of nature. Of course, in ancient thought this idea could only be expressed in crude speculations on the unity of matter, and it soon encountered serious difficulties. By the time of Aristotle it had been abandoned in favour of the concept of a dual universe, comprising the permanent incorruptible realm of the fixed stars and the perishable terrestrial regions where everything suffers change and decay. This presumed duality in nature had to be rejected before modern cosmology could develop. The idea of change had to be extended to the heavens before the principle of the uniformity of nature could be re-established.

The crucial date in the history of this concept was 11th November, 1572. On the evening of that day, Tycho Brahe was contemplating the stars in a clear sky when he noticed a new and unusual star, surpassing the others in brilliancy, shining directly above his head. Observing this star over a period of months, he satisfied himself that it was not a comet but "a star shining in the firmament itself—one that has never previously been seen before our time, in any age since the beginning of the world". Tycho was astounded at his discovery, "For all philosophers agree, and facts clearly prove it to be the case, that in the ethereal regions of the celestial world no change . . . takes place."

After this discovery that change is not confined to the terrestrial regions, the next stage in the development of the idea of the uniformity of nature was dominated by Newton. He succeeded in bringing both terrestrial and celestial phenomena under the sway of the same laws of motion. The subsequent phenomenal growth of both astronomy and physics has been largely due to the application of the idea of uniformity to wider and wider classes of phenomena. In the nineteenth century, Mach suggested that the laws of nature are a direct consequence of the structure of the material universe being what it is. The uniformity of natural law finally came to imply a *basic* uniformity of structure of the material universe. The *a priori* epistemological theories of the modern cosmologists are further developments of this idea.

At the beginning of this chapter a distinction was made between *a priori* methods in the theory of knowledge and in the theory of being. In studying Eddington's philosophy of physics, in particular, it is difficult to escape the impression that he makes certain implicit assumptions about the *existence* of physical objects; nevertheless, both in his theory and in Milne's, fundamental laws of nature, such as the law of inertia and the law of gravitation, are regarded as *ideal* theoretical standards, in terms of which the observed behaviour of bodies in the external world can be most easily interpreted. We are, therefore, justified in regarding the *a priori* foundations of these theories as being mainly epistemological, i.e. concerned with scientific method.

Despite the recent remarkable advances in our theoretical understanding of nature, we are still compelled to admit that "the universe is a hypothesis". Looking out into the depths of space we see the moon a little more than a second in time away, the sun a few minutes, the nearest star a few years, the nearer nebulae hundreds of thousands of years, while the light which we receive from the farthest perceptible nebulae may have been travelling for 500 million years. Judged by terrestrial standards, most of the intervening space is inconceivably empty, but at the limits of our vision there is still no sign of an end. Our idea of the universe as a whole is still a product of the imagination.

"... There was a monk indulging against the teaching of the Master¹ in cosmological enquiries. In order to know where the world ends he began . . . interrogating the gods of the successive heavens. . . . Finally, the Great Brahma himself became manifest, and the monk asked him where the world ends. . . . The Great Brahma took that monk by the arm, led him aside and said: 'These gods, my servants, hold me to be such that there is nothing I cannot see, understand, realize. Therefore, I gave no answer in their presence. But I do not know where the world ends. . . .'"²

¹ *Dialogues of the Buddha.*

² On March 3rd, 1949, it was announced in *The Times* that astronomers on Mount Palomar, after several weeks study of photographs taken there on February 1st with the 200-inch reflector, had concluded that for the first time there had been brought into sight nebulae, appearing only as large as pin-points, 1,000 million light-years away.

SELECT BIBLIOGRAPHY

1. THE OBSERVATIONAL FOUNDATIONS OF COSMOLOGY

An historical approach to the subject can be found in a number of books, including:

W. de Sitter, *Kosmos* (Harvard U.P., 1932), and

H. Macpherson, *Modern Cosmologies* (O.U.P., 1929).

The former is specially recommended.

The recent observational discoveries made with the great telescopes at Mount Wilson and elsewhere are described in:

E. Hubble, *The Realm of the Nebulae* (O.U.P., 1936), and

E. Hubble, *The Observational Approach to Cosmology* (O.U.P., 1937).

The former is technical, the latter semi-popular. Both contain excellent photographs of nebulae. Two lavishly illustrated semi-popular books describing the latest views of the observational astronomers are:

B. J. Bok and P. Bok, *The Milky Way* (Churchill, 1946), and
H. Shapley, *Galaxies* (Churchill, 1947).

The former contains a panoramic photograph of the Milky Way.

An excellent detailed account of the methods described in the first part of Chapter VII is given in:

A. Holmes, *The Age of the Earth* (Nelson, 1937).

2. RELATIVITY AND THEORETICAL COSMOLOGY

A good historical approach to the Theory of Relativity is given in:

A. V. Vasiliev, *Space, Time, Motion* (Chatto & Windus, 1924), and also in:

A. Einstein and L. Infeld, *The Evolution of Physics* (C.U.P., 1938).

Elementary introductions to Relativity include:

E. Borel, *Space and Time* (Blackie, 1931).

A. Einstein, *The Theory of Relativity* (Methuen, 1936).

A. S. Eddington, *Space, Time and Gravitation* (C.U.P., 1920).

A more advanced account, including cosmological applications, can be found in:

E. W. Barnes, *Scientific Theory and Religion* (C.U.P., 1933).

The standard mathematical treatise is:

A. S. Eddington, *The Mathematical Theory of Relativity* (C.U.P., 1922).

The subject matter of Chapter V is treated in great detail in the advanced treatise:

R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (O.U.P., 1934).

A good short account is given in:

G. C. McVittie, *Cosmological Theory* (Methuen's Physical Monographs, 1938), but the last chapter, dealing with Milne's theory, is open to criticism.

A useful book (in German) is:

O. Heckmann, *Theorien der Kosmologie* (Springer Verlag, 1942).

3. EDDINGTON'S THEORY

Two semi-popular books are:

A. S. Eddington, *The Expanding Universe* (C.U.P., 1933), and A. S. Eddington, *New Pathways in Science* (C.U.P., 1935).

Eddington's final views are contained in his advanced treatise: A. S. Eddington, *Fundamental Theory* (C.U.P., 1946).

The philosophical foundations are discussed in:

A. S. Eddington, *The Philosophy of Physical Science* (C.U.P., 1939).

4. MILNE'S THEORY

The two books so far published by Milne on his theory are:

E. A. Milne, *Relativity, Gravitation and World Structure* (O.U.P., 1935).

E. A. Milne, *Kinematic Relativity* (O.U.P., 1948). A semi-popular account is given in:

M. Johnson, *Time, Knowledge and the Nebulae* (Faber, 1945).

The philosophical foundations are discussed by the author in an article on "The Epistemological Foundations of Natural Philosophy", *Philosophy*, Vol. XXI, 1946, pp. 5-28.

5. MISCELLANEOUS LITERATURE

A brief survey of the whole field of modern cosmology is contained in:

J. H. Jeans, *The Astronomical Horizon* (O.U.P., 1946).

A readable discussion of the original and ultimate states of the universe is given in:

E. T. Whittaker, *The Beginning and End of the World* (O.U.P., 1942).

An alternative theory, based on Milne's cosmology, is given in an article by J. B. S. Haldane, "A New Theory of the Past", *The American Scientist*, Vol. XXXIII, 1945, pp. 129-45. The fascinating subject of the evolution of the universe has been almost entirely omitted from the present essay.

Kant's views on space and time are contained in:

Immanuel Kant, *Prolegomena to Every Future Metaphysic* (translated by E. Belfort Bax, George Bell and Sons, 1883), and

Immanuel Kant, *Critique of Pure Reason* (translated by Norman Kemp Smith, Macmillan, 1934).

The relations between cosmology and theology are discussed in:

E. T. Whittaker, *Space and Spirit* (Nelson, 1946).

INDEX

A priori, 103, 148–60
 epistemological, 149, 150, 158, 159
 ontological, 149, 159
 Adams, J. C., 69
 Adams, W. S., 70
 Aldebaran, 11
 Algebra, linear associative, 151
 Andromeda nebula, *see* M 31
 Archimedes, 104
The Sand-Reckoner, 98–100
 Archytas of Tarentum, 44
 Arcturus, 11
 Aristarchus, 18, 99
 Aristotle, 9, 149, 158
 Ayer, A. J., 158

BAADE, W., 129, 136, 145
 Bacon, F., 41
 Barnes, E. W., 74
 Bergson, H., 149
 Berkeley, G., 53, 149
De Motu, 50
 Bessel, F. W., 19
 Bok, B. J., 122, 124, 135
 Boltzmann, L., 77
 Bolyai, J., 54
 Bondi, H., 91
 Bruno, G., 8, 9

CALCIUM, spectral lines, 30, 140, 141
 Cambrian age, 119
 Cambridge Platonists, 46
 Campbell, N., *Physics: The Elements*, 152
 Causality, 155
 Cepheid variables, 22–3, 25, 26, 129, 131, 135, 136
 Chandrasekhar, S., 124

Chapman, S., 96
 Charlier, C. V. L., 76–7
 Clarke, S., 51
 Clausius, R. J. E., 107
 Clifford, W. K., 82
Common Sense of the Exact Sciences, 68
 Comets, tails, 72
 Communicability, principle of, 152
 Constants of nature, 94, 97, 102–5, 106, 114, 151
 Copernicus, N., 9, 21, 39, 40, 51
De Revolutionibus Orbium Caelestium, 8
 Cosmic clouds (*see also* Interstellar matter), 142
 Cosmic rays, 120
 Cosmical constant, 78, 82–5, 88, 95
 Cosmical repulsion, 83, 147
 Cosmological constant (*see* Cosmical constant)
 Creation, 111
 Cygni, 61; 19

DANTE, 116
 Darwin, C. G., 93
 De Sitter, W., 59, 85, 88, 126, 148
 world-model, 81–2, 85, 92, 107
 (with Einstein) model, 83–4
 Delta Cephei, 22
 Delta Orionis, 140
 Descartes, R., 149
 Devonian age, 119
 Digges, T., 8, 9
 Dirac, P. A. M., 93, 95, 102
 Doppler effect (*see also* Red-shifts), 30, 31, 33, 39, 70, 81, 86, 105, 114, 138
 Double stars, 16, 59, 126, 133
 Dunne, J. W., 94

EARTH, age, 116-20, 123, 126
motion, 57-9

Eddington, A. S., 16, 53, 79, 80,
83, 92-105, 109, 120, 122,
139, 141, 150-2, 158, 159

Fundamental Theory, 96, 150

*Relativity Theory of Protons and
Electrons*, 95

uranoid, 97-8, 100-5

Einstein, A. (*see also Relativity*),
56, 59-74, 75, 92, 107, 108,
110
principle of equivalence, 101
universe, 77-81, 82, 83, 85, 86,
88, 94, 95-7, 101, 104, 147,
152
(with De Sitter) model, 83-4

Electromagnetism, 74, 157

Electron, Dirac's theory, 93
Eddington's theory, 94-8, 102-
3, 151
Lorentz's theory, 62

Energy, conservation, 74, 107, 108
equipartition, 121, 122
equivalent to mass, 71-2

Entropy, 108

Eötvös, R., 71

Epicycles, 7

Equivalence (Einstein's principle),
101

Equivalence (Milne's concept),
153-6

Escape, velocity of, 108, 124

Ether, 57-9, 62

Exclusion principle, 97, 104

Euclidean geometry (*see Geometry*)

FARADAY constant for hydrogen,
103

Fine structure constant, 151

Fitzgerald contraction, 59-61, 67,
112

Fizeau's experiment, 62

Formax nebula, 132

Foucault's pendulum, 157

Fraunhofer's lines, 30

Frederick II (Hohenstaufen), 116-
17

Friedmann, A. (*see also Lemaître
universe*), 82, 83, 92

GALAXIES (*see Nebulae, extra-
galactic*)

Galaxy (Milky Way), 11-17, 19-
24, 25, 26, 27, 28, 33, 35, 89,
124, 136-42
absorbing layer, 142-3
age, 121-3, 125, 126
anti-centre, 136, 137
centre, 136-40, 142
dimensions, 23, 28, 134, 135,
136
mass, 24, 125, 140
member of local group, 130-3
poles, 17
rotation, 122, 136, 137-40, 141,
144
shape, 28, 36, 137
stellar distribution, 136-7

Galileo, G., 9, 11, 16, 41, 43,
45-6, 50, 51, 52, 107, 113

Gauss, C. F., 54

Geikie, J., 118

Geometry, 54-5, 154, 155
Euclidean, 12, 46, 54, 78, 82, 83,
84, 86, 104, 108, 111, 112,
154, 155
Lobatchewskian (hyperbolic),
54, 78, 154, 155
non-Euclidean, 12, 54, 67, 77
Riemannian, 68, 74
spherical, 77, 78, 81, 82, 83, 90,
112, 154, 155

George III, 10

Globular clusters, 14, 23, 24, 26,
28, 36, 129, 132, 134, 135,
137, 139

Gravitation, classical (Newtonian),
9, 16, 71, 72, 73, 74, 75, 76,
105, 108, 133, 157
constant of, 78, 96, 98, 101, 103,
114, 157

Gravitation—*cont.*
 relativistic, 65–74, 78, 79, 80,
 81, 83, 87, 92, 94, 95, 147, 159
 velocity of escape, 108, 124
 Gyldén, A., 137

HALLEY, E., 11, 18, 117
 Halm, J., 121
 Hartmann, J., 140, 141
 Harvard College Observatory, 35
 Heckmann, O., 39, 83
 Heisenberg, W., 102, 103
 Helium, 119, 120
 Helmholtz, H. L. F., 118, 155
 Herschel, J., 17, 25, 137
 Herschel, W., 9–17, 19, 20, 21,
 25, 27
 Hipparchus, 11
 Holmes, A., 119–20
 Hoyle, F., 147
 Hubble, E. P., 12, 13, 26, 28, 32–
 40, 104–5, 114, 115, 130, 134,
 135, 142–3, 144, 146
 classification of nebulae, 127–9,
 145
 law, 32–3, 38, 82, 83, 105, 106,
 108
 world-models, 37–9, 86–91
 Huggins, W., 16, 31
 Hurnason, M. L., 32, 89, 90, 91,
 105
 Hume, D., 149
 Huygens, C., 18
 Hyades, 123, 124
 Hydrogen, abundance of, 104, 141
 atom, 97
 Faraday and Rydberg constants,
 103
 spectral lines, 30
 synthesis of elements from, 147
 uranoid, 104

INERTIA, 71, 72, 80, 101, 157
 frame, 65, 80
 law, 65–6, 113, 156–7, 159
 Interferometer, 58

Internebular matter, 88–9
 Interstellar matter (*see also* Cosmic
 clouds), 21–2, 24, 28, 35, 38,
 89, 104, 125, 126, 131, 135,
 136, 140–3
 Isotopes, 119, 120

JEANS, J. H., 23, 121, 125, 139,
 143–4, 146
 Jeffreys, H., 120
 Joly, J., 117
 Jupiter's moons, 52–3

KANT, I., 13–14, 53, 54, 145, 149,
 150
Critique of Pure Reason, 41–2
 Kapteyn, J. C., 20–3, 139
 Kelvin, Lord, 117–18
 Kepler, J., 16
 Kinematical relativity (*see* Milne)
 Kirchhoff, G. R., 30

LAMBDA Herculis, 11
 Lambert, J. H., world-model, 77
 Laplace, P. S., 16
 nebular hypothesis, 145–6
 Lead, 119, 120
 Leavitt, H., 22
 Leibniz, G. W., 51–2, 53, 149, 150
 Lemaître, G., 83, 147
 universe, 83–4, 92, 105, 107
 Leverrier, V. J., 69
 Lick Observatory, 35
 Lie, S., 155
 Light (*see also* Spectra), 17, 21, 37,
 72, 147
 deflection in gravitational field,
 72
 velocity, 19, 52–3, 58–64, 71, 73,
 77, 78, 83, 84, 96, 101, 103,
 106, 109, 110
 Light-year, defined, 19
 Lindblad, B., 125, 137, 139, 143,
 144, 145

Lobatchewskian geometry
 Geometry)
 Locke, J., 149
 Lodge, O., 16
 Lorentz, H. A., 59
 formulae, 63-4, 109, 112
 theory of the electron, 62

M 31; 14, 27, 28, 31, 129, 130, 131, 132, 133, 134, 135, 140, 147
 M 32; 132
 M 33; 133, 134, 145
 M 51; 25
 McCrea, W. H., 90, 108
 Mach, E., 48
 principle, 80, 101, 159
 McVittie, G. C., 39, 90, 91, 132
 Magellanic clouds, 22, 130, 133
 Maupertuis, P. L. M. de, 13
 Maxwell, J. C., 72, 121, 138
 Measurement, theory of, 74, 75, 92, 96, 102, 109, 150-1, 152
 Mercury, 69, 94
 Messier, C., 14
 Meteorites, 120
 Metrology (*see* Measurement, theory of)
 Michelson - Morley experiment, 57-62
 Micro-densitometer, 131, 133
 Milky Way (*see* Galaxy)
 Milne, E. A., 84, 92, 94, 120, 150, 152, 157-8, 159
 equivalence, 153-6
 kinematical relativity, 105-15, 116, 153-6
 substratum, 112-13, 146-7, 156-8
 theory of spiral arms, 144-5
 Minkowski, H. (*see also* Space-time), 64, 67, 73, 93, 94
 Momentum, angular, 147
 linear, 74, 102, 107
 Monads, 149
 Motion, Newton's laws: 1st law (*see* Inertia); 2nd law, 70-1

Mount Palomar Observatory (200-inch reflector), 36, 91, 160
 Mount Wilson Observatory (100-inch reflector), 26, 27, 32, 35, 86, 147

NGC 205; 132
 Nebulae, extra-galactic (*see also* Nebular), 13-17, 19-20, 22, 25-40, 42, 81-9, 127-47, 160
 barred spirals, 128-9, 145
 clusters, 130, 134
 elliptical, 29, 127-32
 irregular, 29, 127, 129-30, 133
 local group, 130-3
 spiral, 25, 26, 28, 29, 31, 32, 123, 127, 128-9, 132, 133, 134, 135, 143-5, 147
 spiral arms, 143-5
 stellar populations I and II, 129, 147
 Nebulae, galactic, 15, 16, 29, 30, 31
 Nebular "colour indices", 89
 dimensions, 27-8, 130-3
 distances, 33, 35, 89, 90, 114
 evolution, 145-6
 magnitude (luminosity), 26-7, 34, 37, 39, 89-90, 114, 115, 133
 mass, 90, 131, 133, 134
 rotation, 134, 143-7
 spectra (*see also* Red-shifts), 30-2, 90, 113-14
 velocities, 31-5, 38, 77, 114, 140
 Neptune, 69
 Newton, I. (*see also* Gravitation, Inertia, Motion, Rotation, Space, Time), 9, 16, 18, 41, 51, 106, 110, 159
Four Letters to Bentley, 75
Principia, 46-50
 static universe, 75-6
 Nicod, J., 55
 Nier, A., 120
 Nuclear fission, 72

OBSERVER, idea of, 52-5, 57-68, 110, 111-12, 152-3
 Olbers, H. W. M., 76-7
 Oort, J., 122, 137, 138, 139
 Orion nebula, 15, 16

PAINLEVÉ, P., 155
 Paneth, F. A., 120
 Parallax, 8, 10, 18, 19, 23
 Parallels, axiom, 54
 Parmenides, 45
 Parsec, defined, 19
 Pascal, B., 7, 9, 148
 Pauli, W., 97, 104
 Pearce, J. A., 141
 Perihelion of Mercury, 69
 Photon, 114-15
 Planck's constant, 102, 103, 106, 114
 Plaskett, J. S., 140, 141
 Plato (*see also* Cambridge Platonists), *Timaeus*, 46
 Pleiades, 124
 Poincaré, H., 54-5, 109
 Pointer-readings, 150
 Protons, Eddington's theory, 95-8, 102-4, 151
 Ptolemaic system, 17
 Pythagoras, 44

QUANTUM theory, 96, 97, 106

RADAR, 109, 110, 112, 153
 Radiation, solar, 89, 118, 125
 Radiation pressure, 72
 Radioactivity, 118, 120, 147
 Radium, 118, 119
 Rayleigh, Lord, 118
 Red-shifts, 30-4, 37, 38, 39, 70, 77, 81, 82, 83, 86, 87, 89, 91, 105, 113-14, 138
 Relativity (*see also* Eddington, Milne), 57-74

General Theory, 65-74, 78, 80, 83-6, 90, 92, 93, 94, 96, 109, 153, 154, 155
 Special Theory, 56, 61-5, 70-4, 82, 85, 93, 109, 113, 152
 Riemannian geometry (*see* Geometry)
 Ritz, W., 59
 Robertson, H. P., 82, 83
 Roemer, O., 52-3, 110
 Rosse, Lord, 25, 143
 Rosseland, S., 141
 Rotation, absolute, 49-51, 80, 95, 156-7
 Ruse, H. S., 90
 Rutherford, Lord, 118
 Rydberg constant for hydrogen, 103

S Doradus, 130
 St. Augustine, 45
 Sagittarius, 23, 136, 139
 Saturn's rings, 138
 Scientific method, 41, 150, 151, 152, 154, 159
 Sculptor nebula, 132
 Seeliger, H., 20, 21, 76, 77
 Shapley, H., 23, 24, 28, 35, 36, 133, 134, 139
 Simultaneity, 60, 95, 107, 110
 Sirius, 11, 18, 31
 companion, 70
 Slipher, V. M., 31, 32, 91, 140
 Soddy, F., 118
 Sodium, 117
 spectral lines, 140, 141
 Space (*see also* Geometry), absolute, 46-7, 50, 53, 57, 62, 64, 79, 156, 157
 and time, 41-56, 59-69, 93, 102
 curvature, 68, 78, 82-6, 87, 88, 90, 108
 dimensions, 151, 157
 Space-time, 64, 67, 68, 73-4, 77, 80, 81, 82, 85, 93, 94, 155

Spectra (*see also* Nebular spectra, Red-shifts, Stellar spectra, Sun), absorption, 29, 30
 continuous, 29
 emission, 29

Spinoza, B., 149, 150

Star clouds, 17, 23, 136

Star clusters (*see also* Globular clusters), 14, 15, 123-5, 126, 129, 137, 139, 142

Stars (*see also* Stellar)
 age, 125-6
 B-type, 121, 122, 129, 133, 142, 145
 Cepheid variables, 22-3, 25, 26, 129, 131, 135, 136
 double, 16, 59, 126, 133
 equipartition of energy, 121, 122
 gauges (counts), 10, 16, 21
 giants, 89-90, 130, 131, 132, 133, 137
 high-velocity, 139
 O-type, 129, 133, 142, 145
 populations I and II, 129, 147
 space-reddening, 142
 streaming, 139
 supernovae, 147, 159

Stebbins, J., 89, 142

Stellar distances, 17, 18, 19, 23
 distribution, 12, 20, 21, 122, 136-7
 energy, 125, 145
 magnitude, 18, 22, 27, 87
 motion, 11, 16, 18, 31, 122
 spectra, 69-70, 140, 141

Strömgren, B., 142

Struve, F. G. W., 19

Substratum (*see* Milne)

Sun, age, 118, 125
 motion, 10, 11, 122, 138, 140
 position, 12, 21, 23, 36, 137
 radiation, 89, 125
 spectrum, 29, 30, 70, 90

Supernovae, 147, 159

TENSOR calculus, 93, 102

Ter Haar, A., 125

Thales, 44, 116, 158

Thermodynamics, laws of, 107-8

Thorium, 119, 120

Time (*see also* Simultaneity), 94, 106-14, 152
 absolute, 46, 47, 60, 62, 79, 81
 dimensions, 151
 irreversibility, 94, 106-8, 153
 proper, 62
 scales (*see also* Universe, age), 109-14, 116, 144, 147, 154, 155
 space and, 41-56, 59-69, 93, 102

Tolman, R. C., 88, 89

Trumpler, R. J., 142

Tycho Brahe, 16, 159

UBBELOHDE, F., *Time and Thermodynamics*, 94

Uncertainty principle, 102, 103

Uniformity principle, 12-14, 16, 34, 152, 154, 158-9

Universe (astronomical), age, 33, 39, 116-26
 density, 35, 90, 104, 146
 expansion, 33-4, 38-9, 105
 extent, 7-40, 87, 104, 160

Universe (theoretical) (*see* World-models)

Uraninite, 119

Uranium, 119, 120

Uranoid (*see* Eddington)

Uranus, 10

VAN de Hulst, A., 125

Vector, 93, 94

Vega, 19

Virgo clusters of nebulae, 29, 134

Vulcan, 69

WALKER, A. G., 90, 154, 155

Weyl, H., 64, 82, 94, 95

Whitehead, A. N., 154

Whitford, A. E., 89

Whittaker, E. T., 90, 104
 Whyte, L. L., *Archimedes or the Future of Physics*, 106
 World-models, 75-115
 de Sitter, 81-2, 85, 92, 107
 Eddington (uranoid), 97-8,
 100-5
 Einstein, 77-81, 82, 83, 85, 86,
 88, 94, 95, 97, 101, 104, 147,
 152
 Einstein and de Sitter, 83-4
 Friedmann (*see* Lemaître)
 Hubble, 37-9, 86-91

Lambert-Charlier, 77
 Lemaître, 83-4, 92, 105, 107
 Milne (substratum), 112-13,
 146-7, 156-8
 Newton (static), 75-6
 Newtonian (expanding), 108
 oscillating, 85, 126
 Wright, T., 12

ZENO's paradoxes, 44, 45
 Zwicky, F., 34

ACKNOWLEDGEMENTS

I wish to express my thanks to F. Booth, of the Department of Theoretical Physics, King's College, London, for his kindness in reading the proofs, and to my wife for her invaluable assistance in preparing the manuscript and compiling the index.

6th June, 1949.

G. J. W.